

Validering av Språk 5-6

- *En valideringsstudie av den nye screeningtesten for språkvansker, Språk 5-6*

Monica Lillevold



Masteroppgave i spesialpedagogikk
Ved det utdanningsvitenskapelige fakultet,
Institutt for spesialpedagogikk

UNIVERSITETET I OSLO

13.11.2009



Sammendrag

Denne oppgaven er et resultat av deltakelse i et større prosjekt som søker å normere og validere førskoleutgaven av screeningtesten Språk 6-16. Den nye testen skal hete Språk 5-6. I denne oppgaven er det sett på sammenhengen mellom Språk 5-6 og spørreskjemaet ”20 spørsmål om språkferdigheter”. Denne sammenhengen ble undersøkt for å kunne si noe om testens validitet. Det er en kvantitativ korrelasjonsstudie hvor til sammen 118 barn deltok. Datamaterialet ble analysert i SPSS.

Oppgavens problemstilling er: *Hvilket forhold er det mellom førskolelæreres vurdering av barns språklige ferdigheter og disse barnas resultater på Språk 5-6?*

For å svare på denne problemstillingen ble to forskningsspørsmål formulert:

1. Måler testen det den skal måle?

Dette spørsmålet handlet om kriterievaliditet. Det ble antatt at testen søkte å si noe om barns språklige ferdigheter. Spørreskjemaet ”20 spørsmål om språkferdigheter” ble brukt som kriterium på dette. Dette spørreskjemaet var allerede validert av Ottem (2009) og kan betraktes som en mer økologisk valid måte å måle språklige ferdigheter på, da informasjonen som innhentes bygger på hverdagsobservasjoner av barna i deres naturlige miljø. Dette minsker risikoen for at deres prestasjoner blir farget av testsituasjonen.

2. Gjør testen nøyaktige målinger?

Dette spørsmålet handlet om spesifisitet og sensitivitet. Altså om testen klarer å plassere de barna med språkvansker i riktig ”bås” og de uten vansker i riktig ”bås”. Dette handler om testens nøyaktighet i vurdering av om barna har vansker eller ikke. For at slutninger trukket fra testresultatene skal være valide, bør testresultatene være nøyaktige. Det bør være minst mulig under- og overrapportering av vansker.

Det ble funnet en signifikant korrelasjon mellom de to instrumentene, som hadde moderat styrke. Dette taler for testens validitet. Analyser av testens nøyaktighet viser at testen er tilstrekkelig nøyaktig i å klassifisere barn uten vansker, men er ikke tilstrekkelig nøyaktig i å klassifisere barn med språkvansker. Dette kan komme av at utvalget inneholdt barn som en antok ikke å ha språkvansker.

Forord

En lærerik og spennende studietid er nå ved veis ende, for denne gang. Jeg sitter nå med min avsluttende oppgave i hendene og i den forbindelse så er det mange som må takkes.

Denne masteroppgaven hadde ikke blitt til uten alle informantene mine. Tusen takk til alle barna som deltok, alle deres foreldre og barnehagene som tok så godt i mot meg.

Tusen takk til mine veiledere Professor Jørgen Frost og spesialpsykolog Ernst Ottem for konstruktiv kritikk og oppfølging underveis i prosessen.

Studietiden ville ikke vært den samme uten alle mine flotte medstudenter. Tusen takk til dere alle, ingen nevnt, ingen glemt. En spesiell takk går derimot til Jeanette, Sigrid og Kristin for korrekturlesing og gode råd.

Jeg har også fått god støtte hjemmefra. Tusen takk til Mamma, Pappa og Maria for all hjelp og støtte.

Til slutt en takk til min kjære Espen som alltid er der for meg.

Oslo, oktober 2009

Monica Lillevold



Innholdsfortegnelse

SAMMENDRAG	III
FORORD	V
1. INNLEDNING	1
1.1 BAGRUNN FOR VALG AV TEMA.....	1
1.2 PROBLEMSTILLING	4
1.3 OPPGAVENS OPPBYGNING	5
2. TEORETISK OG EMPIRISK REFERANSERAMME	7
2.1 SCREENINGTESTER.....	7
2.1.1 Tidspunkt for screening.....	8
2.2 TEORETISK OG EMPIRISK BAKGRUNN FOR SPRÅK 5-6.....	9
2.3 ØKOLOGISK VALIDITET	15
2.4 UTVIKLING I SYNET PÅ VALIDITET	17
2.5 TESTENS NØYAKTIGHET – ET SPØRSMÅL OM SENSITIVITET OG SPESIFISITET.....	21
2.6 EMPIRISK REFERANSERAMME	23
2.6.1 Forskning som benytter de samme instrumentene	23
2.6.2 Forskning om screeningtester versus spørreskjemaer	27
3. METODISK TILNÆRMING	29
3.1 PROSJEKTET	29
3.2 DESIGN OG METODE	30
3.3 UTVALG	32
3.4 PROSEDYRE.....	33
3.5 SPRÅK 5-6.....	34
3.6 SKJEMAET “20 SPØRSMÅL OM SPRÅKFÆRDIGHETER”	39

3.7	ANALYSE	40
3.8	VALIDITET OG RELIABILITET I UNDERSØKELSEN	42
3.9	METODEKRITIKK	43
3.10	ETIKK	44
4.	RESULTATER FRA UNDERSØKELSEN	46
4.1	DESKRIPTIV STATISTIKK	46
4.1.1	<i>Resultater på screeningtesten og spørreskjema.....</i>	<i>46</i>
4.1.2	<i>Reliabilitetsanalyser</i>	<i>52</i>
4.2	ANALYTISK STATISTIKK	55
4.2.1	<i>Analyse av forholdet mellom screeningtesten Språk 5-6 og spørreskjemaet "20 spørsmål om språkferdigheter"</i>	<i>55</i>
4.2.2	<i>Korrelasjonsanalyse</i>	<i>56</i>
4.2.3	<i>Krysstabell</i>	<i>56</i>
4.2.4	<i>Analyser av testens nøyaktighet</i>	<i>58</i>
5.	DISKUSJON	61
5.1	DISKUSJON AV OPPGAVENE I TESTEN	62
5.2	DISKUSJON AV RELIABILITET	64
5.3	DISKUSJON OM TESTENS VALIDITET	67
5.1.1	<i>Vurdering av testens validitet basert på innholdet i testen.</i>	<i>67</i>
5.1.2	<i>Vurdering av testens validitet basert på responsprosess</i>	<i>68</i>
5.1.3	<i>Vurdering av testens validitet basert på testens indre struktur</i>	<i>69</i>
5.1.4	<i>Vurdering av testens validitet basert på dens relasjoner til andre variabler.....</i>	<i>70</i>
5.1.5	<i>Vurdering av testens validitet basert på konsekvenser ved testing.</i>	<i>76</i>
5.4	DISKUSJON AV TESTENS NØYAKTIGHET.	77

5.5 OPPSUMMERENDE VURDERING AV TESTENS VALIDITET	80
6. AVSLUTTENDE KOMMENTARER	81
6.1 VEIEN VIDERE	82
LITTERATUR	84
7. VEDLEGG 1 "20 SPØRSMÅL OM SPRÅKFERDIGHETER"	88
8. VEDLEGG 2: SPRÅK 5 – 6	91
9. VEDLEGG NR 3: INFORMASJONSBREV TIL FORELDRENE	96

Figurer og tabeller

FIGUR 1: MULIGE UTFALL AV EN TEST	22
FIGUR 2: GRAFISK FREMSTILLING AV FORDELINGEN FOR HELE UTVALGET PÅ SPØRRESKJEMAET "20 SPØRSMÅL OM SPRÅKFERDIGHETER"	49
FIGUR 3: GRAFISK FREMSTILLING AV UTFALLET FRA TESTEN OG SPØRRESKJEMAET	60
TABELL 1: RELIABILITETSKOEFFISIENTER SPRÅK 6-16	24
TABELL 2: RELIABILITETSKOEFFISIENTER "20 SPØRSMÅL OM SPRÅKFERDIGHETER"	27
TABELL 3: RESULTATER PÅ SCREENINGTESTEN OG SPØRRESKJEMA	47
TABELL 4: OVERSIKT OVER ITEMS SOM IKKE HADDE VARIANS	53
TABELL 5: RELIABILITET CHRONBACHS ALPHA	54
TABELL 6: KORRELASJONSMATRISSE	56
TABELL 7: KRYSSABELL SPRÅK 5-6 OG 20 SPØRSMÅL OM SPRÅKFERDIGHETER	57

1. Innledning

Dagens samfunn er, i mye større grad enn tidligere, et ”kunnskapssamfunn”. Det har blitt vanligere å ta utdanning ut over videregående opplæring, og arbeidslivet krever mer utdanning enn før (OECD, 1998). Språk er en viktig informasjonskilde samtidig som det er en formidlingskanal. Det å kunne forstå skriftlig og muntlig informasjon er en viktig egenskap for å kunne tilegne seg kunnskap. Samtidig som at skriftlig og muntlig formidlingsevne ofte er viktig i utdanningssammenheng. Slik blir ferdigheter i språk viktig for å lykkes i ”kunnskapssamfunnet”. De som har vansker med språket vil ha et handikap i forhold til å kunne innhente informasjon og formidle kunnskap videre, da en språkvanske kan påvirke både forståelse og evne til å uttrykke seg (Bishop, 1997; Leonard, 1998). Dermed kan en se det slik at det er mer handikappende å ha en språkvanske i et ”kunnskapssamfunn” enn det var på den tiden dagene bestod av å arbeide på gården. Kommunikasjonsmetodene i dagens samfunn har også utviklet seg. En god del kommunikasjon foregår nå i skriftlig form, gjennom e-post, SMS og chatteprogrammer. Det kan gjøre at personer som ikke mestrer den skriftlige formidlingsmåten så godt har vanskeligheter med å delta i denne kommunikasjonen på en fullverdig måte. Slik kan språkvansker føre til at en kan få et handikap på den sosiale arena. Men hva kan gjøres? Å finne de barna som strever med språket så tidlig som mulig, slik at en kan få satt i gang tiltak tidlig, vil være et godt hjelpemiddel. Det at det nå er laget en førskoleutgave av en screeningtest, Språk 6-16, som tidligere er mye brukt i skolen er et skritt på veien mot å kunne finne språkvansker tidligere.

1.1 Bagrunn for valg av tema

Dette prosjektet er et delprosjekt i en større studie: ”Standardisering og validering av en screeningstest av barns språk i alderen 5-6år”, som er et samarbeid mellom Institutt for Spesialpedagogikk (ISP) og Bredtvet kompetansesenter. Det skal utarbeides en ny screeningstest av barns språk, som bygges over samme lest som Språk 6-16. Deltakerne i dette prosjektet skal skaffe normeringsdata ved å teste 5 åringer med denne nye testen. I tillegg skal alle førskolelærerne som har ansvaret for

barna, fylle ut et skjema som kalles ”20 spørsmål om språkferdigheter”, hvor de skal vurdere barnets språklige kompetanse. Det blir også gjennomført en del andre tester som Språk 5-6 skal sammenlignes med for å kunne validere denne. Hovedprosjektets mål er dermed å normere og validere den nye screeningtesten Språk 5-6.

Jeg deltar i dette prosjektet fordi jeg interesserer meg for språk og språkvansker, samtidig som jeg synes det er viktig å fange opp vansker på et tidlig stadium. Derfor var det å kunne få være med på å prøve ut en ny test en mulighet jeg ønsket å benytte meg av. Jeg har valgt å benytte både screeningtesten og spørreskjemaet som data i denne undersøkelsen. Jeg har valgt dette fordi jeg synes det er spennende å kunne se på sammenhengen mellom hvordan førskolelærerne generelt vurderer barnas språklige ferdigheter og hvordan barna presterer på en test. Det er to forskjellige vurderingsgrunnlag. Ved bruk av testing gjøres vurderingen på bakgrunn av en økt hvor barnet skal prestere på ukjente oppgaver administrert av en (i dette tilfellet) ukjent person. Ved bruk av skjemaet har førskolelæreren gjort vurderingen på bakgrunn av hverdagsobservasjoner over lengre tid. Dette delprosjektet vil derfor dreie seg om hvilket forhold det er mellom disse to vurderingsformene, med den hensikt å gjøre vurderinger om validiteten til testen.

Språk 5-6 er en standardisert screeningtest. At en test er standardisert betyr at det finnes prosedyrer som sikrer konsistens i administreringen og skåringen av testen i alle situasjoner (Gall, Gall, & Borg, 2007). Standardiserte tester gir kvantitative data som kan brukes til sammenligninger opp mot fastlagte normer, som er et uttrykk for resultater som er tilegnet under kontrollerte forhold (Marquardt & Gillam, 1999). Spørreskjemaet ”20 spørsmål om språkferdigheter” er ment å være et supplement til tester (Ottem, 2009), og kan være et verktøy for å undersøke hvilke effekter språkvansker fører med seg for barna i kommunikasjon med andre. Når en språkvanske blir til hinder for kommunikasjonen kan det kalles kommunikasjonsvansker (Rygvoid, 2001).

Ved å teste noen med en standardisert test kan en dessuten undersøke hvordan de presterer i forhold til andre. Å se på prestasjonens karakter kan være til hjelp når en

skal bestemme om det dreier seg om en kommunikasjonsvanske eller ikke (Marquardt & Gillam, 1999). Det er imidlertid et viktig poeng at resultater på formelle tester ikke beskriver kommunikasjon i et naturlig miljø. Resultater på tester har sin bakgrunn i svært begrensede utvalg av atferd og skårene vil ikke gi en tilstrekkelig beskrivelse av kommunikativ funksjon hvis de brukes alene (Marquardt & Gillam, 1999). Tester er mye brukt i utredning av språk og talevansker, og har blitt bevist å være valide som verktøy i diagnostisk sammenheng (Marquardt & Gillam, 1999). Testers validitet med tanke på å forutse kommunikative evner i virkelige interaksjoner mellom mennesker har imidlertid ikke blitt tilstrekkelig utforsket (Marquardt & Gillam, 1999).

For å undersøke hvilke følger en språkvanske får for barnet i virkelige interaksjoner med andre kan det være at en må kombinere flere måleinstrumenter. Ved for eksempel å kombinere bruken av Språk 5-6 med "20spørsmål om språkferdigheter" vil en kunne få et mer helhetlig bilde av barnets eventuelle vanske og de implikasjoner denne får for barnet i hverdagen. Hvis barnet har en vanske vil en da ved hjelp av testen se hvordan barnet statistisk sett presterer i forhold til andre barn på samme alder, mens ved hjelp av spørreskjemaet vil en kunne se hvordan barnet fungerer språklig sett i forhold til barn på samme alder, sett med førskolelæreren sine øyne. Ved å undersøke sammenhengen mellom spørreskjemaet og testen kan en finne ut om de språklige områdene som testen undersøker har noen forbindelse med det som førskolelæreren registrerer i hverdagen. En sterk sammenheng mellom de to instrumentene vil være tegn på at testen finner frem til barnets språklige kompetanse slik den viser seg i virkelige situasjoner.

Ligner situasjonen den virkelige verden? Dette spørsmålet omhandler i følge Svartdal (2004) økologisk validitet. Sbordone (1996) definerer økologisk validitet som det funksjonelle og prediktive forholdet mellom prestasjonen på tester og det samme individets atferd i forskjellige reelle settinger (p. 16). På denne måten kan en si at den valideringen som denne oppgaven søker å gjøre av screeningtesten Språk 5-6, er en vurdering av testens økologiske validitet.

1.2 Problemstilling

Problemstillingen for denne oppgaven er:

Hvilket forhold er det mellom førskolelæreres vurdering av barns språklige ferdigheter og disse barnas resultater på Språk 5-6?

Ved å undersøke dette forholdet ønsker en finne ut om resultatene på testen gjenspeiler barnas virkelige språklige ferdigheter slik de fremtrer i reelle situasjoner. Førskolelærernes vurderinger av barnas språklige ferdigheter blir dermed tatt for å være riktige og ”barns språklige kompetanse” blir operasjonalisert som skåren barna får på ”20 spørsmål om språkferdigheter” skjemaet. Da utvalgskriteriene for dette prosjektet utelukket barn med hørselsproblematikk, annet morsmål enn norsk eller barn som var meldt til PPT fra utvalget, er det nærliggende at de språkvansker det vil være snakk om i denne oppgaven er spesifikke språkvansker.

Det finnes en jungel av forskjellige typer validitet, og metoder for vurdering av validitet. I denne oppgaven er det testens kriterievaliditet som vil bli vurdert. Den vil bli vurdert ved å se på sammenhengen mellom to måleinstrumenter, screeningtesten og spørreskjemaet, og i tillegg til kriterievaliditet så vil testens nøyaktighet bli vurdert. Ved å vurdere testen opp mot et mer økologisk valid kriterium vil en kunne vurdere om testen kan fange opp barnas virkelige språklige ferdigheter. Det vil si, slik de fungerer i hverdagen, uhemmet av testsituasjonen og oppgavene.

For å kunne svare på problemstillingen har to forskningsspørsmål blitt benyttet:

1: Måler testen det den skal måle?

Dette spørsmålet tar opp temaet kriterievaliditet. Om en finner en sammenheng mellom screeningtesten og spørreskjemaet vil en kunne svare på om testresultatene samsvarer med et mål av barnas språklige kompetanse i hverdagssituasjoner.

2: Gjør testen nøyaktige målinger?

For at en skal kunne stole på testen som måleinstrument er det viktig at den gjør nøyaktige målinger. Dette vil bli undersøkt ved å regne ut testens sensitivitet og spesifisitet, altså hvor godt den skiller mellom de med og uten språkvansker.

1.3 Oppgavens oppbygning

I kapittel 2 vil det bli gjort rede for teori som belyser problemstillingen og som utgjør en nødvendig forståelsesramme. Temaer som vil bli berørt omhandler screeningtester og hva det er. I og med at dette er en førskoleutgave av en tidligere utgitt test for skolebarn vil også alder tas opp som tema. Testen har en teoretisk og empirisk bakgrunn, hva denne består i vil også bli forklart i kapittel 2. Deretter vil økologisk validitet bli nærmere forklart, og videre en utvikling i synet på validitet i testkonstruksjon. Utviklingen i synet på validitet vil munne ut i hvordan en ser på validitet i dag, og noen måter å vurdere validitet på. Deretter vil det andre forskningsspørsmålet tas opp gjennom spesifisitet og sensitivitet. Til slutt i kapittel 2 vil tidligere forskning som tar opp lignende tema og som benytter de samme instrumentene tas opp.

Kapittel 3 handler om forskningsprosjektets metodiske utforming. Kapitlet vil ta for seg utvalgsmetode og hvilket utvalg vi fikk. Deretter vil de to instrumentene som benyttes i prosjektet bli grundig presentert. En beskrivelse av analysen av data vil også beskrives. Hvilke hensyn og eventuelle problemer det er i forhold til reliabilitet og validitet i dette prosjektet vil også bli berørt. Så vil etiske hensyn som er aktuelle i dette prosjektet bli gjennomgått. Kapitlet avsluttes med noen tanker om hva som kunne vært gjort annerledes.

I kapittel 4 vil resultatene fra denne undersøkelsen presenteres. En gjennomgang av barnas resultater på de to instrumentene vil komme først. Dette gir grunnlag for å kommentere fordelingen form som har betydning for senere statistiske analyser. Så vil instrumentenes indre konsistens analyseres for å kunne si noe om reliabilitet. Instrumentene må ha tilstrekkelig reliabilitet for å kunne benyttes. Deretter vil sammenhengen mellom de to instrumentene undersøkes. En sterk signifikant

sammenheng vil tas til inntekt for testens validitet. Til slutt i dette kapitlet vil testens nøyaktighet bli forsøkt beregnet. Datamaterialet som er tilgjengelig er ikke ideelt for denne beregningen og derfor vil resultatet av denne beregningen bli tolket med forsiktighet.

Kapittel 5 vil inneholde en diskusjon av resultatene som har fremkommet av undersøkelsen. Disse vil bli drøftet i lys av tidligere presentert teori og empiri.

I kapittel 6 vil det presenteres noen avsluttende kommentarer som blant annet vil inneholde noen tanker om hvor veien går videre.

2. Teoretisk og empirisk referanseramme

Formålet med hovedprosjektet er å validere og normere den nye screeningtesten Språk 5-6. Validering av Språk 5-6 foregår ved at den måles opp mot andre tester og spørreskjemaet "20 spørsmål om språkferdigheter". Normeringen gjøres for å kunne vurdere barns prestasjoner opp mot normalen for deres alder.

Ved å sammenligne den nye testen med andre tester kan man få et inntrykk av hvor godt testen måler det den sies å måle. "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests"(AERA, APA, & NCME, 1999, p. 9). Validering er en viktig del av prosessen ved konstruksjon av nye tester. Formålet med dette delprosjektet er å se på sammenhengen mellom screeningtesten Språk 5-6 og spørreskjemaet "20 spørsmål om språkferdigheter". Det gjøres for å se om resultatene fra testen og spørreskjemaet samsvarer og på bakgrunn av dette kan si at testen er et gyldig mål av barnas virkelige språklige ferdigheter.

Screeningtester er en mye brukt metode for innsamling av informasjon om barns kunnskaper og evner, samtidig som de brukes mye i diagnostisering av vansker (Gall, et al., 2007). Positivistisk tankegang handler om at ting skal kunne måles og veies for å få viten om noe (Hjardemaal, 2002). Testing er det nærmeste man kommer slik måling innenfor vårt fagfelt, ettersom det vi studerer ikke er entiteter som man kan ta og føle på, og følgelig heller ikke måle og veie (Hjardemaal, 2002). I dette prosjektet er det språklige ferdigheter som skal undersøkes, med det formål å finne ut om barnet har vansker med språket. Dette kan gjøres ved hjelp av screeningtester.

2.1 Screeningtester

Screeningtester er korte utredninger av et stort antall personer som ikke er dyre å gjennomføre (Thurlow & Gilman, 1999). I dette tilfellet er det snakk om en screeningtest som undersøker barnas språkferdigheter (Ottem & Frost, 2005). Det er

stor enighet om at screeningtester brukes for å sile ut de som ikke trenger å gå igjennom en full utredning (Thurlow & Gilman, 1999). Men det kan også brukes som første steg i en utredning, for eksempel om man er usikker på hvilke vansker det er snakk om.

Screening kan innebære en test, men kan også suppleres med andre former for innhenting av informasjon som observasjon, intervju, utviklingshistorie, sjekklister, og familiefokusert screening og rapporter (Thurlow & Gilman, 1999).

Spørreskjemaet som er benyttet i dette prosjektet kan sies å være en slik annen kilde til informasjon om barnet. For barn som kommer fra familier som tilhører en annen kultur enn majoritetskulturen kan det gi mer nyttig informasjon å bruke slike alternative metoder, ettersom testene kanskje ikke er standardisert med tanke på den spesifikke kulturelle gruppen (Thurlow & Gilman, 1999). Språk 5-6 er laget med tanke på norske barn med norsk som morsmål (Ottem & Frost, 2005). Spørreskjemaet er ikke laget for å besvares av foreldrene, med tanke på at det spør etter sammenligning med andre barn på samme alder, noe det ikke kan forutsettes at foreldre har kunnskap om. Førskolelæreren er noen som kjenner barnet godt og som har kunnskap om barn på samme alder, både fra utdanning og erfaring. Førskolelærerne innehar derimot kanskje ikke den kulturelle kunnskapen som er nødvendig for å vurdere med tanke på kulturelle forskjeller.

2.1.1 Tidspunkt for screening

Det å oppfatte vansker tidlig i forløpet er viktig for å hindre at problemene utvikler seg, og blir mer alvorlige (Knoff, Stollar, Johnson, & Chenneville, 1999). Og med tanke på at ”(...) spesifikke tester ikke fungerer like godt som markører på SSV når de brukes på eldre barn”¹ (Ottem & Lian, 2008, p. 55), kan tidlig screening av barns språk være nyttige i et forebyggingsperspektiv. Språk 5-6 kan være et bidrag til tidlig identifisering av vansker. Tidlig identifisering av vanskene er nødvendig for å få til tidlig intervensjon (Lichtenstein & Ireton, 1984). Det er ønskelig å komme tidlig i

¹ SSV = Spesifikke Språkvansker

gang med hjelp da tidlig intervensjon er forbundet med positiv prognose (Lichtenstein & Ireton, 1984; Rydz, Shevell, Majnemer, & Oskoui, 2005). I teorien vil det å gripe inn tidlig i vanskenes forløp, før de utvikler seg og blir store vansker, være en god bruk av ressurser. Da kan man bruke ressursene på ”ounces of prevention” i stedet for ”pounds of cure” (Lichtenstein & Ireton, 1984, p. 6). Det er derimot viktig å ikke starte for tidlig med å lete etter vansker. Barn er i stor utvikling i førskolealder, og ikke alle utvikler seg i samme tempo. Derfor kan det være vanskelig å si hva som er en vanske og hva som ikke er det. For tidlig identifisering av vansker kan føre til overrapportering og unødig fokusering på vansker. Det er ikke sikkert at alle de som tilsynelatende hadde vansker på et tidlig tidspunkt, vil vise seg å ha vansker senere på grunn av ulikt tempo i utviklingen (Lichtenstein & Ireton, 1984). Likevel må en screene tidlig, men en slik screening må ikke føre til stempling av barna. Screeninga må være grunnlag for tiltak som skal kunne bedre situasjonen for de barna som strever. Barnehagen er en arena hvor det er rom for forskjellighet, og generelle tiltak kan rettes mot gruppen som helhet.

Ved at den nye screeningtesten Språk 5- 6 nå er laget, og at den bygger på en test som allerede er innarbeidet i skolen (Språk 6-16), har man nå muligheten til å oppdage tegn på språkvansker tidligere i forløpet enn man kunne ved bruk av Språk 6-16. Samtidig kan resultater fra Språk 5- 6 videreformidles fra barnehagen til skolen. Da vil lærerne kanskje allerede være kjent med oppbygningen av testen og har kanskje selv brukt Språk 6-16 og vet hvordan denne fungerer. Det vil kanskje føre til at kommunikasjonen mellom barnehage og skole vil være enklere og kvalitativt bedre.

2.2 Teoretisk og empirisk bakgrunn for Språk 5-6

Screeningtesten Språk 5-6 er bygget en antakelse om at det er en sammenheng mellom begrensninger i verbalt korttidsminne og barns vokabularutvikling (Ottem & Frost, 2005, p. 6). Denne antakelsen bygger på funn gjort av Gathercole og Baddeley (1990) og Baddeley, Gathercole og Papagno (1998). Før disse forskernes funn blir

presentert vil Baddeley og Hitches (1974) arbeidsminnemodelle presenteres for å gi et forståelsesgrunnlag.

Baddeley og Hitch (1974) sin modell bygger på at de hadde en formening om at menneskets korttidsminne var bygget opp som et arbeidsminne, som bestod av tre deler. En sentral eksekutiv enhet, og to slavesystemer: Den fonologiske løkken og den visuospatiale skisseblokken. Senere har Baddeley (2000) lagt til enda et slavesystem: Den episodiske bufferen. Videre vil det bli gitt en kort forklaring av de forskjellige elementene i modellen for arbeidsminnet.

Den sentrale eksekutive enhet:

Denne enheten antas å være et oppmerksomhetssystem med begrenset kapasitet. Den koordinerer aktivitet i arbeidsminnet og kontrollerer utvekslingen av informasjon med andre deler av det kognitive systemet (Gathercole & Baddeley, 1993). Enheten kontrollerer også de tre slavesystemene: Den fonologiske løkken, den visuospatiale skisseblokken og den episodiske bufferen.

Den fonologiske løkke:

Løkken som helhet tar seg av fonologisk informasjon (Alan Baddeley & Hitch, 1974; Gathercole & Baddeley, 1993). Den fonologiske løkken består av to deler: et fonologisk korttidslager med et auditivt minnespor. Det auditive minnesporet forfaller raskt og vil forsvinne uten den artikulatoriske øvelseskomponenten som gjenoppliver minnesporene (Alan Baddeley, 2000; Gathercole & Baddeley, 1993). Slik går informasjonen som i en løkke, den kommer inn i det fonologiske korttidsminnet og gjennom artikulatorisk øvelse blir den gjentatt og minnet blir gjenoppfrisket.

All auditiv verbal informasjon antas å gå direkte inn i det fonologiske lageret. Språk som blir presentert visuelt kan gjøres om til fonologisk kode ved stille artikulering og slik bli kodet inn i det fonologiske minnet (Gathercole & Baddeley, 1993). Om objekter som kan navngis blir presentert visuelt vil navnet på disse kunne overføres fra den visuospatiale skisseblokken til den fonologiske løkken ved at objektene navngis subvokalt (Alan Baddeley, 2002). Den fonologiske løkken kan spille en

viktig rolle i tilegnelsen av vokabular, spesielt i tidlig barneår (Alan Baddeley, et al., 1998). Ved studier av verbalt korttidsminne blir fonologisk minne oftest målt ved at testpersonen må gjenta en liste av items, dette kan være kjente ord eller tall (Alan Baddeley, 2000; Hulme & Roodenrys, 1995).

Screeningtesten Språk 5-6 undersøker det fonologiske korttidsminnet gjennom deltesten ordspenn. Her er det lister med kjente ord som barnet skal gjenta. Deltesten setningsminne er også en minneoppgave, men da ordene kan relateres til hverandre og skape mening kan barnet dra nytte av langtidsminnet. Slik vil ikke dette være en ren test av det fonologiske minnet da barnet også kan dra nytte av langtidsminnet i denne oppgaven.

Det er vanskelig å bevise hvordan ting henger sammen når det gjelder funksjoner i hjernen, men forskning har kommet frem til fem hovedfunn som styrker teorien om den fonologiske løkken:

1. Fonologisk likhetseffekt

Forskning har vist at det er mye vanskeligere å huske ordrekker hvor ordene rimer på hverandre enn det er å huske rekker med distinkte ord (Alan Baddeley, 1966). Dette kalles fonologisk likhetseffekt. Ord som rimer er mer lik hverandre enn distinkte ord, slik at det er færre enheter som skiller dem fra hverandre. Om det for eksempel er én lyd som skiller ordene fra hverandre (pil – bil), og barnet glemmer den første lyden i ordet, er det ikke mulig å huske rekkefølgen. Derimot hvis det er distinkte ord er det flere enheter i ordet som skiller dem fra hverandre, da vil ikke det å glemme første bokstav ha like store konsekvenser (Gathercole & Baddeley, 1993).

I deltesten ordspenn er annenhver ordrekke satt sammen av ord som rimer og annenhver av distinkte ord. På den måten økes vanskelighetsgraden og den fonologiske likhetseffekten blir undersøkt.

2. Ordlengdeeffekten

Baddeley, Thomason og Buchanan (1975) fant at minnet for kortere ord er bedre enn minnet for lengre ord, dette kalles ordlengdeeffekten. Disse funnene tyder på at minnet har begrensninger i tid (Alan Baddeley, et al., 1975). Ved at lengre ord tar lengre tid å si, og dermed lengre tid å øve på, har minnene av de foregående ordene lengre tid på seg til å forsvinne (Alan Baddeley, et al., 1975). Både på deltesten setningsminne og deltesten ordspenn økes lengden på setningene og ordremsene etter hvert. Slik får oppgavene en økende vanskelighetsgrad.

3. Effekten av artikulatorisk undertrykkelse

Artikulatorisk undertrykkelse skjer ved at testpersonen blir bedt om å artikulere irrelevant materiale, som for eksempel "at", samtidig som lister med ord eller tall skal gjengis korrekt (Alan Baddeley, et al., 1975; Gathercole & Baddeley, 1993). Den artikulatoriske undertrykkelsen gjør det umulig for testpersonen å øve subvokalt på ordene som skal gjentas (Alan Baddeley, et al., 1975; Gathercole & Baddeley, 1993). En får da frem et renere mål på det fonologiske minnet ved at en hindrer effekten av den subvokale øvelsesprosessen.

4. Overføring av informasjon mellom koder

Når gjenstander blir presentert visuelt, vil voksne testpersoner vanligvis gi navn til gjenstandene og øve på ordene subvokalt, slik at informasjonen blir overført fra en visuell til en auditiv kode (Alan Baddeley, 2000).

Artikulatorisk undertrykkelse forhindrer denne overføringen, og på den måten vil den tidligere nevnte effekten av fonologisk likhet, bli visket ut for visuelt presenterte items (Murray, 1968). Screeningtesten Språk 5-6 inneholder ikke visuelt presentert materiale.

5. Nevropsykologiske beviser:

Atferden til pasienter med en spesiell svikt i fonologisk korttidsminne kan forklares av et defekt fonologisk lager. Et eksempel er afatiske pasienter med

dyspraksi. Disse har ikke muligheter til å organisere talemotoriske koder som er nødvendige for å kunne artikulere, på grunn av en svikt i den artikulatoriske øvelsesprosessen (Waters, Rochon, & Caplan, 1992)

Derimot har pasienter med dysartri², hvis talevansker er sekundære, normal kapasitet for øvelse. Dette tyder på at det er den subvokale øvelsen som er viktig (Alan Baddeley & Wilson, 1985).

Den visuospatiale skisseblokken:

Den visuospatiale skisseblokken er et slavesystem som prosesserer og lagrer visuell og spatial informasjon (Alan Baddeley & Hitch, 1974). Slik som det å huske former og farger, eller et objekts hastighet eller plassering i rommet. Dette slavesystemet spiller ikke en stor rolle for språket (Gathercole & Baddeley, 1993) og vil dermed ikke bli utdypet nærmere her.

Den episodiske bufferen:

I 2000 la Baddeley til en fjerde komponent i modellen, den kalles episodisk buffer (Alan Baddeley, 2000). Denne komponenten er et tredje slavesystem, som tar seg av å sette sammen informasjon mellom de andre slavesystemene for å danne integrerte pakker av visuell, spatial og verbal informasjon i kronologisk orden, slik som et minne av en historie eller en scene fra en film (Alan Baddeley, 2000). Den episodiske bufferen antas også å ha kontakt med langtidsmminnet og semantisk betydning (Alan Baddeley, 2000; Dewey, 1995).

Kanskje kan deltesten setningsminne undersøke funksjonen i den episodiske bufferen. For å løse setningsminnetesten kan det være nødvendig å gjøre nytte av informasjon fra langtidsmminnet i form av sammenhenger og benytte arbeidsminnet til å holde på informasjonen lenge nok til å kunne gjengi den.

² Dysartri er en talemotorisk vanske som kommer av nedsatt styrke, hurtighet eller koordinasjon av talemuskulaturen (Darley, Aronson, & Brown, 1975).

Funnene som danner bakgrunn for testen

Gathercole og Baddeley(1990) fant at barn med forstyrret språklig utvikling var dårligere til å repetere nonord og å gjenkalle ordlister enn kontrollgruppene. De fant også at barna med forstyrret språkutvikling var påvirket av fonologisk likhetseffekt og ordlengdeeffekt på samme måte som barn med normal språkutvikling (Gathercole & Baddeley, 1990). Videre eksperimenter førte til en konklusjon om at den dårlige prestasjonen på minnetestene av barna med språkvansker ikke var på grunn av svekket perseptuell prosessering eller langsom artikulasjon (Gathercole & Baddeley, 1990). De kom frem til et forslag om at en svikt i det fonologiske lageret i arbeidsminnet kan være en underliggende årsak til at barna med språkvansker gjorde det dårlig på minnetestene og at den mangelen kan være forstyrrende i deres språkutvikling (Gathercole & Baddeley, 1990). Baddeley, Gathercole og Papagno (1998) studerte ett av slavesystemene i Baddeley og Hitches (1974) modell av arbeidsminnet, den fonologiske løkken. Gjennom deres studie av undersøkelser av normale voksne og barns, nevropsykologiske pasienters, og spesifikke utviklingspopulasjoner ordlæring fant de bevis for at den fonologiske løkken er en viktig brikke i det å lære nye fonologiske former av nye ord (Alan Baddeley, et al., 1998). På bakgrunn av sine funn foreslår forskerne at den fonologiske løkkens primære oppgave er å lagre ukjente lydmønstre mens mer permanente minner blir konstruert (Alan Baddeley, et al., 1998). Den fonologiske løkken blir dermed viktig i vokabularutviklingen, i møtet med ukjente ord.

Screeningtesten Språk 5-6 bygger på en antakelse om at barn med språkvansker har en svikt i det fonologiske lageret. Denne utforskes av deltestene setningsminne og ordspenn. En slik svikt kan skape problemer med å lære nye ord, noe som vil vise seg i barnets vokabular. Det blir undersøkt gjennom deltesten begreper. Screeningtesten undersøker altså fonologisk minne og begrepsmessig utvikling (Ottem & Frost, 2005).

Informasjonen fra spørreskjemaet er basert på en mer helhetlig vurdering av barnets semantiske evne, evne til å formidle og motta budskap og barnets relasjoner til andre. En helhetlig vurdering står sterkt innen et økologisk perspektiv på kartlegging.

2.3 Økologisk validitet

Nå har den teoretiske og empiriske bakgrunnen for testen blitt presentert. Den viste at testen bygger på visse antakelser om hvilken rolle det fonologiske minnet spiller i barnets begrepsmessige utvikling og at en svikt i det fonologiske minnet kan ligge til grunn for en forstyrret språklig utvikling. For å undersøke om en kan si at screeningtesten Språk 5-6 er et godt mål av barns språklige ferdigheter kan en se på sammenhengen mellom denne og spørreskjemaet ”20 spørsmål om språkferdigheter”. Informasjonen fra spørreskjemaene er innhentet i naturlige situasjoner og kan på den måten sies å gjenspeile den virkelige verden. Som nevnt i innledningen handler økologisk validitet om det kan sies at situasjonen ligner den virkelige verden (Svartdal, 2004). I dette delkapitlet skal begrepet økologisk validitet og dens aktualitet for dette prosjektet utdypes.

Å ha et økologisk perspektiv på det å samle informasjon om barnet kan være spesielt viktig med tanke på sosiale og emosjonelle vansker, eller vansker i relasjoner til andre (Knoff, et al., 1999). Sosiale og emosjonelle vansker kan være nært knyttet til språkvansker (Botting & Conti-Ramsden, 2000). Spørreskjemaet ”20 spørsmål om språkferdigheter” inneholder, i tillegg til de 20 spørsmålene, også 7 spørsmål om barnets relasjoner til andre. Disse belyser et aspekt som både kan ses i sammenheng med språkvansker, og som også kan opptre alene.

Atferd skjer ofte i samspill med andre og bør derfor ikke vurderes isolert. Derfor kan det være lurt å vurdere barnet i sitt naturlige miljø i samspill med andre (Knoff, et al., 1999). Det bør også tas i betraktning at problemer med atferd kan være et uttrykk for andre vansker. Det er ikke bare atferd som behøver en vurdering som innebefatter miljøet rundt. En kan overføre denne tenkningen til å gjelde utredning av språk også, ved å legge et kommunikasjonsperspektiv på språket. Språk er en svært viktig del av

kommunikasjon, og kommunikasjon foregår i samspill med andre. Derfor er det viktig at språk også ses i sammenheng med hvordan eventuelle vansker påvirker kommunikasjonen i barnets naturlige miljø. Hundeide (2008) foreslår at "(...) den tradisjonelle testbaserte, individualistiske sviktdiagnosen bør suppleres med en intersubjektiv diagnose eller en kartlegging av den meningen denne svikten har i barnets sosiale virkelighet" (pp. 185-186). Det Hundeide (2008) her skriver peker i retning av en mer økologisk tenkning rundt utrening av barns vansker.

I følge Bronfenbrenner (1979) kan økologisk validitet defineres slik: "Ecological validity refers to the extent to which the environment experienced by the subjects in a scientific investigation has the properties it is supposed or assumed to have by the investigator" (Bronfenbrenner, 1979, p. 29). Bronfenbrenner fokuserer på den oppfatningen barnet har av kartleggingssituasjonen og at den skal være den samme som den intenderte situasjonen fra testlederens side. Han mener at tenkningen rundt økologisk validitet bringer med seg et viktig aspekt, nemlig at de miljømessige kontekstene som en undersøkelse blir gjort i kan påvirke prosessen i kartleggingssituasjonen og dermed påvirke tolkningen og generaliserbarheten til funnene (Bronfenbrenner, 1979). Slik kan miljøet påvirke resultatenes gyldighet.

Enhver utredning med tanke på et barns eventuelle språkvansker, har som mål å finne ut av barnet sterke og svake sider. Barnets profil kan være et godt verktøy når en skal planlegge tiltak. Økologisk validitet er viktig med tanke på tiltak (Marquardt & Gillam, 1999). Tester er nyttige verktøy for beskrivelse av vansker, men brukt alene er de ikke tilstrekkelige for å planlegging av tiltak, eller for å beskrive hvilke konsekvenser vanskene får for barnet (Marquardt & Gillam, 1999, p. 265). En skal bare utføre en screeningtest for å komme i forkant av en eventuell vanske hvis dette er til det beste for barnet og samfunnet (Rydz, et al., 2005, p. 9). Det beste for både barnet og samfunnet vil være å finne frem til eventuelle vansker så tidlig som mulig med sikte på å avhjelpe disse. For å kunne stole på slutninger om barnets vansker må testen ha en viss gyldighet.

2.4 Utvikling i synet på validitet

Denne delen skal handle om hva som er særegent med validitet i testing. Det er en forskjell på de validitetsvurderinger man gjør fra et testkonstruksjonsstandpunkt og den validitetsvurderingen man gjør av en undersøkelse. Undersøkelsens validitet handler om hvordan ting ble gjort, designets svakheter og styrker i forhold til validitet (Befring, 2007; Shadish, Cook, & Campbell, 2002). I testkonstruksjon handler det om å konstruere et verktøy som er så nøyaktig som mulig og som man kan trekke gyldige slutninger fra (AERA, et al., 1999; Gregory, 2000; Murphy & Davidshofer, 2001).

Noen amerikanske organisasjoner, som befatter seg med testing innenfor fagfeltene psykologi, oppvekst og utdanning, har satt sammen en offisiell håndbok om standarder innen testing. De mener disse standardene bør overholdes for at testingen skal være gyldig og etisk forsvarlig (Gall, et al., 2007). Det er blitt publisert flere utgaver av denne håndboken som kalles: *Standards for Educational and Psychological Testing* (heretter referert til som *Standardene*). Gjennom snart 60 år har en kunnet følge utviklingen i synet på validitet gjennom disse standardene. Det vil nå bli gitt et kort sammendrag av denne utviklingen for å kunne gi en bedre bakgrunnsforståelse.

Det har skjedd en forandring i synet på validitet i tester. På 1940-tallet handlet validitet om testen i seg selv (Goodwin & Leech, 2003). Det var testen som skulle valideres og denne kunne da omtales som å inneha validitet. Validitet ble sett på som en statisk egenskap ved en test, som ble bestemt av korrelasjon med et ytre kriterium (Goodwin & Leech, 2003). Da standardene for 1966 kom ut hadde fokuset skiftet fra korrelasjon til bruk (Goodwin & Leech, 2003). Testen ble betraktet som valid hvis den kunne være nyttig for et spesielt formål (Goodwin & Leech, 2003). 1966-utgaven av standardene introduserte også en tredeling av validitet; innholdsvaliditet, kriterievaliditet og begrepsvaliditet. Dette var egentlig ikke noe nytt, ettersom denne inndelingen først ble presentert av Cronbach og Meehl i 1955 (Cronbach & Meehl, 1955). På 1980-tallet skiftet vindene i synet på validitet retning igjen. Det ble da satt søkelys på de slutningene som tas på bakgrunn av testskårene (Goodwin & Leech,

2003). Testvalidering ble beskrevet som prosessen med å samle bevis for å støtte slike slutninger (Goodwin & Leech, 2003). På 1980 og 1990 tallet ble nytten av tredelingen av validitet utfordret. 1985-utgaven av standardene inkluderte også tredelingen av validitet, men påpekte at det ikke skulle regnes som ulike typer validitet (Goodwin & Leech, 2003). Validitet begynte å bli sett på som et enhetlig konsept, hvor begrepsvaliditet var den viktigste og samlende faktoren. Noe som også ble mer aktuelt i 1980 og 1990-årene, var vurdering av sosiale konsekvenser ved bruk av tester (Goodwin & Leech, 2003). Det nye i denne sammenhengen var å se på både planlagte og uplanlagte konsekvenser av testingen.

I den foreløpig siste utgaven av ”Standardene” blir validitet sett som en enhet, og det som tidligere ble betegnet som forskjellige typer validitet, blir nå betegnet som tegn på validitet (AERA, et al., 1999; Goodwin & Leech, 2003). I denne utgaven av Standardene blir konsekvenser av testing introdusert som et tegn på validitet:

The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations. It is the interpretations of test scores required by proposed uses that are evaluated, not the test itself. When test scores are used or interpreted in more than one way, each intended interpretation must be validated (AERA, et al., 1999, p. 9).

Den gamle tredelingen av validitet blir erstattet av en ny inndeling som fokuserer på fem forskjellige typer av validitetsvurdering; vurdering basert på *indre struktur*, vurdering av *testinnhold*, vurdering basert på *responsprosess*, vurdering basert på *forhold til andre variabler* og vurdering basert på *konsekvensene av testing* (AERA, et al., 1999; Goodwin & Leech, 2003). Det skal videre bli gitt en kort presentasjon av disse typene vurdering av validitet.

Vurdering basert på innholdet i testen

Dette er spørsmål om hvorvidt innholdet i testen representerer domenet³ til det fenomenet som skal studeres (AERA, et al., 1999; Goodwin & Leech, 2003). Her skal man vurdere testens tilstrekkelighet, klarhet, relevans, og hvorvidt det er samsvar mellom de enkelte items og oppgaver i testen, og definisjonen av begrepet (AERA, et al., 1999; Goodwin & Leech, 2003). I forhold til den testen som skal vurderes i dette prosjektet, vil det si at man for eksempel undersøker om de minneoppgavene som er tatt med, og de ordrekkene som skal huskes samsvarer med definisjonen av fonologisk minne og hvordan dette undersøkes. Her tas også i betraktning ”construct-irrelevant variance” og ”construct underrepresentation”, som handler om hvorvidt testen synes å måle mer eller mindre hva den var ment å teste (AERA, et al., 1999; Goodwin & Leech, 2003).

Vurdering basert på responsprosessen

For denne typen vurdering må man undersøke om oppgavene, eller typen responser som kreves av de som blir testet, passer til det begrepet som skal undersøkes (AERA, et al., 1999; Goodwin & Leech, 2003). Her ser man på om oppgavene i Språk 5-6 er egnet til å undersøke de underliggende faktorene som de er ment å måle, som for eksempel fonologisk minne.

Vurdering basert på indre struktur

Ved en vurdering basert på indre struktur undersøkes det om de interne komponentene i en test stemmer overens med det definerte begrepet (AERA, et al., 1999; Goodwin & Leech, 2003). I denne sammenhengen vil det være å se om fonologisk minne og begrepsmessig utvikling, som er de underliggende komponentene som Språk 5-6 sies å måle, er viktige faktorer med tanke på språkvansker.

³ En hver psykologisk test undersøker et utdrag av atferd fra et domene (Murphy & Davidshofer, 2001). Domenet i dette tilfellet er språklige ferdigheter. Screeningtesten Språk 5-6 undersøker et utdrag av atferd som tilhører domenet språklige ferdigheter, fonologisk minne og begrepsmessig utvikling (Ottem & Frost, 2005) for å avdekke eventuelle svikter i disse ferdighetene, noe som kan tyde på språkvansker (Gathercole & Baddeley, 1990).

Vurdering basert på relasjoner til andre variabler

Vurdering basert på relasjoner til andre variabler er den største kategorien. Noen av de mest kjente metodene for å samle inn slik informasjon er korrelasjonsstudier og eksperimentelle studier (Goodwin & Leech, 2003). Denne formen for vurdering av tegn på validitet er mest aktuell for dette prosjektet da det er en korrelasjonsstudie. Korrelasjonsstudier ser blant annet på hvordan, og i hvilken utstrekning, det er sammenheng mellom skårer og eksterne kriterievariabler (tilegnet på samme tidspunkt eller senere) (Goodwin & Leech, 2003). Det er også studier av forholdet mellom testskårer og data innsamlet med andre instrumenter eller måleteknikker som antas å være et mål av det samme begrepet og studier av forholdet mellom skårer og mål på begreper med forskjellig meningsinnhold (Goodwin & Leech, 2003). Når testen sammenlignes med et mål som antas å måle det samme begrepet, er det konvergent validitet som undersøkes (Gregory, 2000). Hvis testen sammenlignes med et mål som måler noe helt annet, som denne testen normalt sett ikke skal korrelere med, er det diskriminerende validitet som vurderes (Gregory, 2000). I dette prosjektet vil en se på sammenhengen mellom skårene på Språk 5-6 og en ekstern kriterievariabel, som er ”20 spørsmål om språkferdigheter” skjemaet. Data fra spørreskjemaet er tilegnet på omtrent samme tid som data fra testen.

Vurdering av konsekvenser ved testing

Denne typen vurdering av tegn på validitet vedrører både ventede og uventede konsekvenser av målingen (AERA, et al., 1999; Goodwin & Leech, 2003). Tester brukes vanligvis i den hensikt at det vil komme noe godt ut av det (AERA, et al., 1999). “A fundamental purpose of validation is to indicate whether these specific benefits are likely to be realized” (AERA, et al., 1999, p. 16). Cronbach (1988) understreket betydningen og viktigheten av å opparbeide et validitetsargument og inkludere vurderingen av konsekvenser i forsøk på å validere. I følge Cronbach (1988) inneholder et slikt argument alle bevisene for og imot foreslåtte tolkninger av skårer, inkludert eventuelle vurderinger om rimelige alternative tolkninger. Mulige konsekvenser av det å teste barn for å se om de har språkvansker vil være å finne ut at

de har eller ikke har vansker. Ved tilstedeværelsen av eventuelle vansker vil barnet kunne få ekstra hjelp og støtte. Konsekvensene må også ses i sammenheng med testens nøyaktighet. Om en benytter en test som er unøyaktig er det fare for overrapportering eller underrapportering. Dette kan få konsekvenser i form av feilaktig identifisering av vansker eller at vansker blir oversett.

Slik situasjonen er nå ses ikke validitet på som noe en kan oppnå 100 % og som varer for alltid, det ses derimot på som en kontinuerlig prosess. Det er slutningene fra resultatene på testen som må være valide. Og validiteten til slutningene må vurderes på nytt om de skal brukes på en annen måte enn det som var planlagt. Det er fem typer validitetsvurderinger som kan tas i bruk: Vurdering av validitet basert på innholdet i testen, vurdering basert på responsprosessen, vurdering av validitet basert på testens indre struktur, vurdering basert på relasjoner til andre variabler og vurdering av konsekvenser ved testing. Som tidligere nevnt vil det bli sett på kriterievaliditet, altså vurdering basert på relasjoner til andre variabler, i dette prosjektet. Det vil også fokuseres på testens egnethet til å gjøre nøyaktige målinger.

2.5 Testens nøyaktighet – Et spørsmål om sensitivitet og spesifisitet.

For at en skal kunne gjøre valide slutninger fra testresultater, må målingene være nøyaktige (Murphy & Davidshofer, 2001). For at en test skal være nøyaktig må utfallene fra den være så riktige som mulig. Man kan beskrive en screeningtests suksess ved å dele inn utfallene av testen i fire grupper (Rydz, et al., 2005):

1: Hvis testen identifiserer noen uten vansker som positiv (med vansker), kalles det falsk - positiv, som fører til en overrapportering av vanskene.

2: Hvis testen identifiserer noen som har vansker som positiv (med vansker), da er denne sann-positiv.

3: Hvis testen identifiserer noen som har vansker som negativ (uten vansker), kalles det falsk- negativ, dette fører til underrapportering av vansker.

4: Hvis testen identifiserer noen uten vansker som negativ (uten vansker), kalles det sann-negativt.

Figur 1 - Mulige utfall av en test

Sann positiv (2)	Falsk negativ (3)
Falsk positiv (1)	Sann negativ (4)

(Murphy & Davidshofer, 2001, p. 183)

Figur 1 viser de mulige utfallene av testingen. For at man skal kunne stole på testen som bakgrunn for å ta avgjørelser er det viktig at man reduserer forekomsten av de to utfallene, falsk-negativ og falsk-positiv. Nøyaktigheten til testen blir vurdert gjennom *sensitivitet* og *spesifisitet* (Rydz, et al., 2005). Sensitiviteten sier noe om sjansen for å identifisere alle individer med språkvansker (Rydz, et al., 2005). Spesifisiteten handler om sjansen for å korrekt identifisere noen som ikke har vansker (Rydz, et al., 2005). Hvis testen har 100 % sensitivitet vil den finne alle *med språkvansker* i utvalget, og hvis den har 100 % spesifisitet vil screeningtesten utelukke alle som *ikke har språkvansker* i utvalget, fra språkvanskegruppen. Dess høyere prosent av sensitivitet og spesifisitet testen har, dess mindre sjanse er det for underrapportering og overrapportering (Rydz, et al., 2005). Det kan beskrives som en sil, man må ha store nok hull i silen til at de uten språkvansker slipper igjennom, men ikke så store at de som har språkvansker også faller igjennom.

For å kunne sette dette prosjektet i en større sammenheng og kanskje sammenligne med andre studier vil det nå bli gitt en kort gjennomgang av relaterte studier.

2.6 Empirisk referanseramme

I 1997 skrev Dorothy Bishop :” To my knowledge, nobody has tried to relate statistical test-based diagnoses of SLI to ratings by parents, teachers, peers, or even children themselves, of how far communication problems affect everyday life” (Bishop, 1997, p. 26). Det vil i dette delprosjektet bli forsøkt å gjøre noe lignende det som Bishop etterspør her. Det er ikke brukt en testbasert diagnose av spesifikke språkvansker, men en test som screener for språkvansker og det er førskolelærere som svarer på spørsmål om hvordan de opplever at vanskene påvirker barnet i hverdagen. Ut fra Bishops utsagn kan en anta at det ikke finnes store mengder forskning på dette området, men noe er det. Det vil nå bli gitt en kort innføring i noen av disse.

2.6.1 Forskning som benytter de samme instrumentene

Screeningtesten Språk 5-6 er førskoleutgaven av Språk 6-16 som ble utgitt i 2005. Testen ble utprøvd på 1080 barn og ungdom fordelt på fire geografiske regioner i Norge: Øst, vest, sør og nord (Ottem & Frost, 2005). Barn som hadde språkvansker eller ikke hadde norsk som det dominerende språket ble ekskludert fra undersøkelsen (Ottem & Frost, 2005). Det var ønskelig med barn som representerte normal språklig utvikling for å kunne bruke data fra prosjektet til sammenligning. Det ble satt en kritisk grense på -1,5 standardavvik under gjennomsnittet, og tall fra denne undersøkelsen viste at 6-7 % av elevene befant seg under denne kritiske grensen (Ottem & Frost, 2005). Det ble blant annet vurdert kriterievaliditeten til testen hvor kriteriet som testen ble målt opp mot var lærernes vurdering av elevenes språklige kompetanse på en skala fra 1 til 5. Ottem og Frost (2005) fant at det var høyt samsvar mellom lærernes vurdering og testresultatene ($\text{Chi-square} = 223,97$, $\text{df} = 16$, $p < .000$) (p. 25). Testens innholdsvaliditet ble også vurdert. Denne ble vurdert ved å se på hvor godt testen skilte mellom barn som var diagnostisert med språkvansker og barn med antatt normal språkutvikling. 89 % av barna med påviste språkvansker ble riktig klassifisert av testen, og 84,7 % av barna med antatt normal språkutvikling ble riktig klassifisert (Ottem & Frost, 2005, p. 27). Feilprosenten på 15,3 i gruppen med antatt

normal språkutvikling har sammenheng med at i denne gruppen fantes det også barn med språkvansker (Ottem & Frost, 2005). Testens reliabilitet ble undersøkt ved hjelp av å se på indre konsistens, og ble beregnet ved Cronbachs Alpha.

Reliabilitetskoeffisienter finnes også fra en senere studie da det skulle lages nye normer for å gjøre det lettere å sammenligne testresultatene fra Språk 6-16 med andre tester (Ottem, 2007). For den ble utvalget utvidet til 1214 personer.

Reliabilitetskoeffisientene fra begge undersøkelsene er presentert i tabellen nedenfor:

Tabell 1 Reliabilitetskoeffisienter Språk 6-16

Språk 6-16	Setningsminne	Ordspenn	Begreper	Sum
2005	.81	.75	.90	.93
2007	.85	.71	.90	.93
Antall items	16	12	26	54

(Ottem, 2007; Ottem & Frost, 2005)

En ser her at tallene fra 2007 er ganske like de fra 2005, men med en liten stigning i reliabilitetskoeffisienten for setningsminne og en litt lavere koeffisient for ordspenn. Alle reliabilitetskoeffisientene unntatt de for ordspenn er over grensen på .80 som kreves for at instrumentet kan brukes i forskningssammenheng (Gall, et al., 2007). Ottem og Frost (2005) forklarer den lave reliabilitetskoeffisienten for ordspenn med at det skyldes testens korthet og at tester som måler minnespenn generelt sett har lavere reliabilitetskoeffisienter enn kunnskapstester.

Kriterievaliditet ble undersøkt for dette utvalget også, på samme måte som i 2005 ved å sette testskårene opp mot lærernes vurdering av elevenes språklige kompetanse, vurdert på en skala fra 1 til 5 (Ottem, 2007). Det ble også undersøkt om de nye normene kunne bidra til å identifisere barn med språkvansker. Testens nøyaktighet i å skille normalbarn fra barn som var henvist Bredtvet kompetansesenter for språkvansker var 88 %, basert på deltestene setningsminne, ordspenn og begreper. I 2005 ble som tidligere nevnt testens nøyaktighet til å klassifisere barna med språkvansker riktig beregnet til 89 %.

I arbeidet med å standardisere og normere Språk 6-16, og i den forbindelse valideringen i form av å se testen opp mot lærernes vurdering av elevenes språklige kompetanse på en skala fra 1 til 5, fant man en sterk sammenheng, men også tilfeller hvor lærerne og testen var uenige (Ottem, 2009). Dette forholdet var det derfor ønskelig å undersøke i mer detalj. Det ble utviklet et spørreskjema som bestod av en rekke utsagn om språkvansker. Det skjemaet som ble utviklet er det samme som ble benyttet i dette prosjektet (en utførlig beskrivelse av spørreskjemaet som ble brukt i dette prosjektet finnes i kapittel 3.6). Spørreskjemaet bestod av 20 utsagn om språklige problemer som lærerne skulle gradere etter hvor godt de passet for den aktuelle eleven, fra 1 til 5. De 20 utsagnene var gruppert i tre delskalaer: semantisk evne, vansker med å motta et budskap og vansker med å formidle et budskap. I tillegg ble det konstruert en skala som omhandlet barnets ”relasjoner til andre”, den kommer utenom de 20 spørsmålene og inneholder 7 utsagn. Denne skalaen ble tatt med fordi det er funnet sterke sammenhenger mellom språkvansker og psykisk helse (Se for eksempel: Beitchman, et al., 2001; Botting & Conti-Ramsden, 2000). ”Relasjoner til andre” ble tatt med som en representasjon av ett aspekt ved sammenhengen mellom språkvansker og psykisk helse (Ottem, 2009).

Ved hjelp av en strukturell regresjonsanalyse basert på data fra de typiske barna fant Ottem (2009) at ”kunnskap om ords betydning (”semantisk evne”) ligger til grunn for både det å oppfatte hva andre sier og selv å kunne uttrykke seg slik at en blir forstått av andre” (Ottem, 2009, p. 16). Ottem (2009) fant også lave regresjonskoeffisienter mellom de enkelte delskalaene og alder, og koeffisientene representerte heller ingen signifikante sammenhenger, en kan derfor si at delskalaene er aldersuavhengige for typiske barn (Christophersen, 2006; Ottem, 2009).

Spørreskjemaet ble utprøvd på 496 elever i normalskolen og elever som var henvist til Bredtvet kompetansesenter for vansker med språket (Ottem, 2009). Det var lærerne til elevene som svarte spørreskjemaene. Det ble foretatt en analyse av sammenhengen mellom spørreskjemaet og Språk 6-16. Denne analysen er gjort på

bakgrunn av data fra 93 typiske barn og 98 barn med forskjellige typer vansker, data fra to-språklige barn ble utelukket (Ottem, 2009).

Et spørsmål i skjemaet handler om barnet hovedvanske og på bakgrunn av svarene på dette spørsmålet ble de elevene som hadde andre vansker enn språkvansker utelukket. Slik endte de opp med et utvalg av 250 typiske barn og 48 barn med språkvansker. På bakgrunn av data fra disse 298 elevene fant Ottem (2009) at 31 % var en kritisk grense. En skåre over 31 % av maksimal skåre forekommer relativt sjelden hos typiske barn og ettersom poengsummen stiger forekommer den skåren sjeldnere og sjeldnere hos typiske barn. Kun ca 13 % av de typiske barna har skårer over 31 % av maksimal skåre (Ottem, 2009). Ettersom utvalget også inneholdt barn med språkvansker, var det mulig å undersøke hvordan denne kritiske grensen passet for barna med språkvansker. Ved å sette kritisk grense til 31 % av maksimal skåre så fant Ottem at skårene til ca 83 % av elevene med språkvansker befant seg over denne kritiske grensen (Ottem, 2009).

For å undersøke spørreskjemaets validitet ble det undersøkt om det var en sammenheng med screeningtesten Språk 6-16. Utvalget av elever hvor det var benyttet både spørreskjemaet og Språk 6-16 var på 191 elever, av disse var 93 typiske elever og 98 var elever med antatte vansker (Ottem, 2009). Ottem (2009) fant en klar sammenheng mellom skjemaet og testen ($\text{Chi-square} = 83.6$, $\text{df} = 16$, $p < .001$) (Ottem, 2009, p. 17). Selv om det var en klar sammenheng så var ikke resultatene entydige, en nærmere undersøkelse av materialet viste derimot at forskjellene kunne forklares ut fra elevenes profilmønster. Et eksempel var :Eleven fikk toppskårer på to av de obligatoriske deltestene i Språk 6-16, mens han fikk bunnskårer på den tredje obligatoriske deltesten (Ottem, 2009). En slik skåre kan føre til at eleven skårer relativt høyt sammenlagt, men at han har betydelige vansker på en av deltestene. Det er sannsynligvis denne vansken som læreren oppfatter og gir uttrykk for i sin skåring av elevens språklige ferdigheter (Ottem, 2009).

Reliabilitet måler spørreskjemaets indre konsistens og ble i dette tilfellet beregnet ved Cronbachs alpha (Ottem, 2009). For beregningen av reliabilitet bestod utvalget av

elever med språkvansker og typiske barn (Ottem, 2009). Reliabilitetskoeffisientene for 20 spørsmål og delskalaen ”relasjoner til andre”, ble som følger:

Tabell 2 Reliabilitetskoeffisienter ”20 spørsmål om språkferdigheter”

Reliabilitet	α
20 spørsmål	.97
Relasjoner til andre	.88

En ser av tabell 2 at både delskalaen ”relasjoner til andre” og skjemaet ”20 spørsmål om språkferdigheter” har svært god reliabilitet. Begge koeffisientene befinner seg godt over grensen på .8.

2.6.2 Forskning om screeningtester versus spørreskjemaer

En studie av Fletcher og Satz (1984) viste at tester predikerer bedre de som er i risikosonen for å utvikle lesevansker enn hva lærere gjør. Lærerne i deres studie viste seg å ha høye falske negative verdier. Selv om lærerne hadde lave falsk-positive verdier (8%), fant de bare 41% av de sann-positive. Det vil si at de overså 59 % av de som virkelig hadde eller kom til å få vansker med lesing (Fletcher & Satz, 1984). Deres konklusjon ble at, når det er viktig å finne de som har vansker er tester bedre prediktorer enn lærere (Fletcher & Satz, 1984). Black (1979) mener derimot at en viktig innvending mot standardiserte tester er deres innebygde føringer for kommunikasjon. I en mer naturlig setting ville barna kunne bruke sitt eget vokabular og talemåter og gi uttrykk for deres livsforståelse, deres assosiasjoner og verdier (Black, 1979). Patton (1975 i Black, 1979) er enig med Black og mener at standardiserte tester ikke er noen erstatning for førskolelærerens nøye gjennomtenkte vurderinger om barn som de kjenner godt.

Det finnes flere måter å kartlegge språkvansker på. Man kan se etter språklige ferdigheter generelt og se om noen ligger dårligere an enn det som er normalt for alderen. Eller man kan se etter risikofaktorer for språkvansker. Flynn og Rahbar

(1998) ville i sitt forskningsprosjekt forsøke å forbedre førskolelærerprediksjoner av barn som ville slite med *leseinnlæringen*, ved å utvikle et måleinstrument som hadde god innholds- og prediktiv validitet. Et av deres formål med prosjektet var å avgjøre om screeningbatteriet de brukte gav nok prediktiv nytte sammenlignet med førskolelærervurderingene for å rettferdiggjøre kostnadene ved bruken av det (Flynn & Rahbar, 1998). De fant lav sann-positiv rate når førskolelærere skulle si hvordan de trodde det ville gå med leseutviklingen til barna på en skala fra 1 til 5 (Flynn & Rahbar, 1998). Førskolelærerne rangerte barna på skalaen. De barna som førskolelærerne gav lav skåre (1 og 2 poeng), som indikerer vansker, fanget bare opp 30 % av barna som hadde lesevansker i 3. klasse. Derimot fanget screeningbatteriet opp 81 % av barna som fikk lesevansker innen 3. klasse (Flynn & Rahbar, 1998). Flynn og Rahbar (1998) laget derfor en ny skala som skulle korrespondere med deltestene i screeningbatteriet. De fant ut at det kanskje ville være bedre å be førskolelærerne om å uttale seg om barnas nåværende evner i sammenheng med indikatorer på leseevne. De trodde også at det å kombinere resultater fra en screeningtest med førskolelærervurderinger kanskje ville forbedre prediksjonsratene (Flynn & Rahbar, 1998). De fant at førskolelærervurderinger ofte gir et mer helhetlig og annerledes bilde av barns preleksikale evner sammenlignet med resultatene på testen som ble brukt (Flynn & Rahbar, 1998). Det skjemaet som brukes i dette prosjektet, ”20 spørsmål om språkferdigheter”, er laget på en lignende måte. Førskolelærerne skal uttale seg om deres vurdering av barnet, det vil kunne bygge på en vurdering av utvikling, men vil for det meste dreie seg om hvilke språkferdigheter barnet innehar nå. De utsagnene førskolelærerne skal ta stilling til om passer for barnet, er utsagn som passer til tegn på språkvansker (Ottem, 2009).

3. Metodisk tilnærming

For at en undersøkelse skal kunne være etterprøvbar er det viktig å beskrive metoden for innsamling og analyse av data. Det vil i dette kapitlet derfor bli gjort rede for metodisk tilnærming i hovedprosjektet og for dette delprosjektet. Design, metode og instrumenter for datainnsamling vil bli presentert. Deretter vil det bli gjort rede for analysene som er gjort av datamaterialet. Til slutt vil reliabilitet, validitet og etiske betraktninger bli berørt.

3.1 Prosjektet

Hovedprosjektet har som mål å bidra til normeringen og valideringen av Språk 5-6. Bidraget til normeringen skjer ved at studentene samler inn data fra barn i aldersgruppen 5 til 6 år ved å teste normalfungerende barn i denne aldersgruppen. Bidraget til valideringen skjer ved at studentene i tillegg til screeningtesten Språk 5-6 har utført en del andre tester på barna som skal brukes til sammenligning med Språk 5-6. Gruppen med studenter ble delt i to, og delte testene som skulle brukes til sammenligning mellom seg. Slik at ikke alle barna i utvalget har blitt testet med de samme testene. Derimot har *alle* studentene brukt den nye testen, Språk 5-6, og *alle* har benyttet spørreskjemaet ”20 spørsmål om språkferdigheter”.

En fordel ved å være en del av et forskningsprosjekt er at man kan benytte seg av data som andre i prosjektet har samlet inn. I dette forskningsprosjektet deltar 5 studenter som forskningsassistenter. Dermed har dette delprosjektet fått et større utvalg enn en hadde klart å få på egenhånd. Dette er en fordel med tanke på generalisering av resultater til populasjonen. Studentene gikk ut med den intensjonen å teste 25 barn hver, det ville tilsi et utvalg på 125 barn. Vi har fått et samlet utvalg på 118 barn, av forskjellige årsaker, blant andre sykdom og barn som var bortreist. N er dermed lik 118 i dette delprosjektet ettersom alle studentene har administrert de måleinstrumentene som er aktuelle i dette delprosjektet.

Et kvantitativt forskningsopplegg skal være etterprøvbart. For å oppnå en slik etterprøvbarhet må en beskrive hvilken design som ligger til grunn for undersøkelsen, og hvilke metoder som er benyttet. Det må gjøres rede for utvalgsmetode og en beskrivelse av utvalget en fikk, dette er også for at man skal kunne kritisere forskningen på grunn av eventuelt skjevt utvalg. Fremgangsmåte vil bli beskrevet i delkapitlet 3.4, her beskrives både forberedelsene *før* forskningen ble påbegynt, og hvordan testingen og innsamlingen av informasjon gjennom skjemaene ble gjennomført. Deretter kommer en mer utførlig beskrivelse av de instrumentene som ble brukt, hvor skåringsregler og hvordan de brukes blir forklart.

3.2 Design og metode

En av de vanligste måtene for forskere å samle informasjon om individer er å administrere tester (Gall, Gall, & Borg, 1996). Som nevnt gir tester kvantitative data. Problemstillingen søker å si noe om sammenhengen mellom to variabler. For å kunne si noe om en sammenheng, må man se etter korrelasjoner. Dette delprosjektet har dermed et korrelasjonsdesign. Barna er kartlagt på bare ett tidspunkt, men med to forskjellige metoder, screeningtest av språk og spørreskjemaet ”20 spørsmål om språkferdigheter”. Undersøkelsen har et ikke-eksperimentelt design. Et slikt design karakteriseres av at det ikke søker å forandre noe, det skal beskrive hvordan ting er og hvordan variablene forholder seg til hverandre (Kleven, 2002a). Det er vanskelig å si noe om årsaksforhold i en undersøkelse med et ikke-eksperimentelt design og undersøkelsen vil derfor ha lavere indre validitet⁴ enn ekte eksperimentelle design (Kleven, 2002a). I dette delprosjektet vil ikke fokus ligge på hvorfor noe er som det er, men på om det er en sammenheng mellom de to variablene, ”20 spørsmål om språkferdigheter” og screeningtesten Språk 5-6. Tester er en god måte å undersøke et individs ferdigheter i noe, som for eksempel språk (Robson, 2002). For å finne ut av hvordan noen vurderer noe, kan spørreskjema være en god metode (Robson, 2002).

⁴ Indre validitet handler om gyldigheten av slutninger om årsaksforhold mellom variablene (Kleven, 2002a). Ettersom en ikke kan uttale seg om årsaksforhold ved hjelp av de variablene og det designet en har valgt, så vil denne studien ha lavere indre validitet enn for eksempel en studie med eksperimentelt design.

Gjennom det spørreskjemaet som anvendes i dette prosjektet søker vi å finne ut av førskolelærernes vurderinger av barnas språklige ferdigheter.

Spørreskjemaet er utformet som et skjema med 20 utsagn. Førskolelærerne skal ta stilling til hvor godt utsagnene passer til det enkelte barn. Førskolelærerne vurderer hvert utsagn på en skala fra 1 til 5, altså en Likert skala (Robson, 2002). Det er ment å bli fylt ut uten at førskolelærerne går inn for å observere barna. Dette fordi man ønsker å få tak i førskolelærernes generelle inntrykk av barnas språklige ferdigheter. Et slikt spørreskjema er mindre ressurskrevende for førskolelærerne å gjennomføre, i forhold til en systematisk observasjon av hvert enkelt barns språklige ferdigheter.

Det å finne ut hvordan en skal omtale det skjemaet som her er benyttet, har vært vanskelig. Ottem (2009) kaller det et observasjonsskjema, noe han begrunner med at informasjonen vi får fra det er grunnet i hverdagsobservasjoner foretatt av førskolelærerne. Jeg har derimot et inntrykk av at det må ligge en intensjon bak observasjonene for at det skal kalles observasjon (Se for eksempel: Vedeler, 2000). Samtidig som det i dette tilfellet ikke ville være forskeren som foretok observasjonen, men førskolelærerne (Befring, 2007). Mye kan observeres i hverdagen, men for at det skal bli en del av den grenen av forskningen som kan kalles observasjonsforskning, må observasjonen være tilsiktet (Robson, 2002; Vedeler, 2000). Det kan også kalles spørreskjema ettersom tittelen på skjemaet er ”20 spørsmål om språkferdigheter”. Skjemaet inneholder derimot ingen spørsmål, bare utsagn som forskolelærerne skal ta stilling til hvor godt de passer til det aktuelle barnet. Én type spørreskjema som kalles ”inventory”(Lichtenstein & Ireton, 1984), kan sies å være noe tilsvarende det skjemaet som er brukt. Jeg velger å oversette ”inventory” med liste. Dette hentyder til at skjemaet består av en liste med trekk som tyder på språk- og kommunikasjonsvansker, samt vanskeligheter i relasjoner til andre.”20 spørsmål om språkferdigheter” blir derfor omtalt som liste eller spørreskjema/skjema heretter.

Lister er strukturerte, selvadministrerende spørreskjemaer hvor informantene beskriver seg selv eller andre (Lichtenstein & Ireton, 1984). I dette prosjektet brukes ”20 spørsmål om språkferdigheter”, som er en standardisert liste hvor førskolelæreren

beskriver barnet. Lister inneholder utsagn (items) og begrensede svaralternativer (Lichtenstein & Ireton, 1984). Skjemaet som er brukt har svaralternativer på en skala fra 1 til 5. Items blir ofte gruppert for å skape skalaer som måler forskjellige områder eller aspekt ved individets fungering (Lichtenstein & Ireton, 1984). ”20 spørsmål om språkferdigheter” inneholder tre skalaer av grupperte items som beskriver henholdsvis ”semantisk evne”, ”evne til å motta et budskap” og ”evne til å formidle et budskap”. Disse skalaene gir en global oversikt over barnas språklige ferdigheter (Ottem, 2009). Skalaen ”semantiske evner” beskriver rene språkferdigheter, mens ”vansker med å motta/formidle et budskap” har et mer kommunikativt perspektiv (Ottem, 2009). I tillegg til skalaene som beskriver språklige ferdigheter inneholder listen en skala som sier noe om barnets ”relasjoner til andre”. Ved å inkludere denne skalaen kan det undersøkes om barnet viser tegn på atferdsvansker. Disse kan knyttes til den eventuelle språkvansken, eller kan ses som tegn på en selvstendig vanske.

Lister har ofte normer som resultatene kan tolkes opp mot (Lichtenstein & Ireton, 1984). Den listen som brukes i dette prosjektet, er ikke prøvd ut på barn mellom 5 og 6 år. Dermed har en ikke noen normer å forholde seg til, men analyser av ”20 spørsmål om språkferdigheter” viser at den er aldersuavhengig (Ottem, 2009). Lister er selvadministrerende og dermed kostnadseffektive. Resultatene kan sammenfattes på en systematisk måte, enten i en sumskåre eller en profil (Lichtenstein & Ireton, 1984). Listen ”20 spørsmål om språkferdigheter” vil bli beskrevet i kapittel 3.6. En viktig faktor for å bedømme et forskningsstudie er det utvalget som ble benyttet. Utvalget i dette prosjektet vil nå bli beskrevet.

3.3 Utvalg

Utvalget i dette prosjektet er skaffet til veie av de fem studentene som er med i prosjektet. Vi gikk ut med like kriterier for utvelgelse, og hadde forfattet brevene til både foreldre og barnehagene sammen. Slik sikret vi at alle fikk den samme informasjonen. Kriteriene barna skulle velges ut fra, var som følger: Barna skulle være mellom 5,0 og 5,11,30 år, de skulle ikke ha hørselsproblematikk, de skulle ha

norsk som morsmål og de skulle ikke være meldt til PPT. Dette var på grunn av at vår oppgave i hovedprosjektet var å skaffe normaldata til normeringen av testen. Vi trengte derfor barn som man antok ikke hadde vansker. Det utvalget vi fikk består av 118 barn, mellom 5 og 6 år. Ett barn i undersøkelsen fylte seks år den dagen testingen ble foretatt, dette barnet er likevel tatt med i dette delprosjektet. Vi har testet 64 jenter og 54 gutter. Det vil si at utvalget er ca 53% jenter og ca 46% gutter. 27 av 118 barn kommer fra en middels stor kommune på vestlandet, det utgjør ca 23 %, resten av barna kommer fra østlandet. Selv om noen av barna har en mor eller en far som har et annet morsmål enn norsk, så har alle oppgitt at barnets morsmål er norsk, og alle har oppgitt at barnets hjemmespråk er norsk. Ingen barn har blitt utelukket fra utvalget som ligger til grunn for denne oppgaven fordi en trenger en viss variasjon av barn for å kunne bestemme testens validitet. Testen er konstruert for å finne barn som har vansker med språket, da kan det være nyttig å prøve den ut på barn som ikke er ytterligere selektert.

For enkelhetsskyld omtales alle barna i denne oppgaven som hankjønn. Dette gjøres for å kunne skille dem fra førskolelærerne og testlederne som omtales som hunkjønn. Det kan kanskje sies at førskolelærerne som deltar i denne studien også er informanter, men de gir ikke informasjon om seg selv, og de anses derfor ikke som informanter i den forstand.

For å kunne gjenskape en studie må man vite hvordan forskningen ble gjennomført (Hjardemaal, 2002). Det er det jeg nå skal forklare.

3.4 Prosedyre

I dette prosjektet hadde vi behov for å bruke barn som informanter. Barn har redusert samtykkekompetanse og vi måtte derfor be foreldrene om tillatelse til å utføre screening av barnet deres, samtidig som de gav samtykke til at førskolelæreren kunne gi informasjon om barnet gjennom spørreskjemaet (NESH, 2006).

Derfor sendte vi ut et informasjonsbrev med svarslipp til foreldrene i god tid før vi skulle ut i barnehagene. I dette brevet informerte vi om prosjektet og hvilke tester vi ønsket å gjennomføre og hva de skulle brukes til. Vi satte en frist for å levere tilbake svarslippene, for de som ønsket å være med, på ca. en uke. Slik sikret vi oss å få inn svarene tidsnok til at alle tillatelser var i orden til vi skulle starte. Ved å sette fristen en uke frem i tid ønsket vi å gi foreldrene god nok tid til å tenke over saken. Samtidig ville vi ikke at fristen skulle være for langt frem i tid slik at vi kunne risikere at det kunne bli glemt. I informasjonsbrevene ble det forsøkt å gi så klar og fyldig informasjon som mulig. Dette ble gjort for å sikre et informert samtykke fra foreldrene. Det innebærer at foreldrene må ha nok informasjon til å forstå hva det er de gir sitt samtykke til (NESH, 2006). Det ble gjort klart at det var frivillig å være med, og at det var lov til å trekke seg på et hvilket som helst tidspunkt uten at dette ville få noen følger for dem eller barnet (se vedlegg nr 3). Slik ønsket en å sikre at foreldrenes samtykke var fritt, det vil si at de ikke følte seg presset til å være med (NESH, 2006). Selv om foreldrene hadde gitt sitt samtykke til at barnet deres kunne delta i studien var vi forberedt på å avbryte testing, eller unnlate å teste barn som vegret seg for å være med. Det ble derimot ikke noe problem ettersom de fleste barna gledet seg over å delta.

Testingen av barna foregikk individuelt i barnas barnehage. Vi fikk låne et rom i barnehagen hvor testingen fant sted. Screeningtesten Språk 5-6 tar ca 20 min å gjennomføre og ble da planlagt tatt i ett strekk uten pauser. Førskolelærerne fikk beholde ”20 spørsmål om språkferdigheter” skjemaene og fylle dem ut når de hadde tid til det.

Videre vil testprosedyren og prosedyren for å fylle ut listen ”20 spørsmål om språkferdigheter” bli forklart.

3.5 Språk 5-6

Språk 5-6 er navnet på en ny språkscreeningtest som er laget av Ernst Ottem og Jørgen Frost. Denne er bygget over samme lest som Språk 6-16. På samme måte som

sin ”storebror” er testen konstruert for å kunne fange opp tegn på språkvansker, men i motsetning til Språk 6-16, som er for skolebarn, er Språk 5-6 laget for barn som går siste året i barnehagen. Testen består av de tre obligatoriske deltestene fra Språk 6-16: *Setningsminne*, *ordspenn* og *begreper*. Begreper er en todelt oppgave og består av: *Motsetninger* og *ordforklaring*. Forskningsprosjektet som dette delprosjektet er en del av er i ferd med å normere og validere den nye testen. Det foreligger foreløpig ingen manual til den nye testen, men ettersom den er svært lik Språk 6-16, med unntak av noen enklere items som er lagt til hver av oppgavene, har vi brukt manualen tilhørende Språk 6-16 i gjennomføringen av testingen. Hvilke svar som skulle godtas på begrepsoppgaven har vi diskutert i prosjektgruppen.

Setningsminne

Deltesten Setningsminne undersøker barns evne til å gjenta setninger. Konfirmatorisk faktoranalyse peker i retning av at Setningsminne kan sees på som en funksjon av fonologisk minne og begrepsmessig utvikling (Ottem & Frost, 2005, pp. 27 - 29). Oppgaven går ut på at testleder leser en setning høyt, som barnet skal gjenta. Testleder introduserer deltesten med å si at hun skal lese noen setninger høyt og barnet skal gjenta akkurat det samme som hun sier. Det blir først gitt en prøveoppgave for å se om oppgaven er forstått. Det er viktig å forsikre seg om at barnet har forstått oppgaven før testingen begynner, for å være sikker på at man får undersøkt de underliggende funksjonene som oppgaven er ment å teste, og for å sikre at barnet ikke får dårligere skåre på grunn av at han ikke har forstått oppgaven. Da dette er en minneoppgave, er det ikke tillatt med gjentakelser. Prøveoppgaven er: *Kari har ei grønn kåpe* (Ottem & Frost, 2005).

Når barnet har forstått prøveoppgaven og gjentatt den riktige setningen, går man videre til selve gjennomføringen av deltesten. Forskjellen på Språk 5-6 og Språk 6-16 er her at det har blitt lagt til fire enklere setninger i begynnelsen. Det vil si at barn mellom 5 og 6 år begynner med to-ords-setninger (*Gutten løper*). Setningene blir lengre etter hvert. De enkleste setningene inneholder to ord, de vanskeligste inneholder 13 ord. Setningene leses i normalt tale tempo (Ottem & Frost, 2005).

Det er tillatt for testleder å tilpasse setningene til barnets dialekt. For eksempel kan det være nødvendig å bytte ut ordet *løper* i *gutten løper*, med *springer*. Å springe er et mer anvendt verb enn *løpe* i enkelte deler av landet. Testleder står fritt til å ”oversette” setningene etter beste evne for å imøtekomme barnets dialekt. Dette kan gjøres fordi det ikke er barnets bokmålkunnskaper man skal teste, men barnets språkferdigheter, i dette tilfellet fonologisk minne og begrepsmessig utvikling. Hvis barnet blir bedt om å gjenta noe som faller ham svært unaturlig å si, kan det forstyrre konsentrasjonen og dermed gi et feil inntrykk av barnets fonologiske minne. Han kan henge seg opp i ord eller uttrykksmåter som kan bli en feilkilde i testingen, og da i den testskåren barnet sitter igjen med.

Det gis 1 poeng for riktig gjentatt setning, og 0 poeng for én eller flere feil (Ottem & Frost, 2005). Maksimal skåre på denne deltesten er 20 poeng. Testingen skal avbrytes etter 3 påfølgende feil.

Ordspenn

Deltesten *ordspenn* er ment å måle barnets *fonologiske minne* (Ottem & Frost, 2005). Testleder introduserer deltesten ved å si at hun skal lese noen ord og at barnet må gjenta de samme ordene i samme rekkefølge. Også her blir barnet presentert for en prøveoppgave for å undersøke om han har forstått oppgaven. Prøveoppgaven er: *Tann-Sekk-Kokk*. Ordene leses med en fart på ca. ett ord i sekundet (Ottem & Frost, 2005).

I denne oppgaven har det, som i setningsminne, blitt lagt til noen enklere items i begynnelsen. Barn på 5-6 år begynner med ordrekker på to ord, den vanskeligste ordrekken inneholder fem ord. Det er ikke tillatt med gjentakelser, fordi man vil undersøke barnets fonologiske minne, noe som krever at barnet ikke har mulighet til å ”lære” eller huske fra sist gang han hørte den samme ordrekken. Det kan sammenlignes med test-retest-effekt hvor skåren på retesten kan bli kunstig høy fordi barnet husker fra sist gang han ble testet (De Vaus, 2002).

Annenhver ordrekke består av ord som rimer (*Ball- Kall- Fall*), og annenhver består av distinkte ord (*Rusk- Hegg - Fikk*). Dette gjøres fordi forskning viser at det er vanskeligere å huske ord som rimer på hverandre, også kalt fonologisk likhetseffekt (Alan Baddeley, 1966; Gathercole & Baddeley, 1990). Denne effekten påvirker både typiske barn og barn som har vansker med språket (Alan Baddeley, 1966; Gathercole & Baddeley, 1990). Det er lettere for barna, når ordene skal gjengis i riktig rekkefølge, å blande sammen ordene når de rimer på hverandre. Ved å inkludere rekker med ord som rimer og ikke bare distinkte ord, blir oppgavene dermed mer utfordrende.

Det gis 1 poeng for riktig gjengitte ord i riktig rekkefølge, og 0 poeng for en eller flere feil (Ottem & Frost, 2005). Det tas hensyn til at barn kan ha artikulasjonsfeil, og testleder skal være oppmerksom på disse, og gi poeng hvis ordet er gjentatt riktig, men med feil uttale. Maksimal poengsum på denne deltesten er 16 poeng. Stoppkriteriet er at testleder skal stoppe etter tre påfølgende feil (Ottem & Frost, 2005).

Begreper

Deltesten *begreper* er ment å måle *begrepsmessig utvikling* (Ottem & Frost, 2005). Den er todelt og omfatter både kunnskap om synonymer og om antonymer.

Motsetninger

Her skal barnet finne det motsatte ordet av stimulusordet, altså antonymet. Denne oppgaven er behold slik den er i Språk 6-16. De tre første oppgavene er presentert i en setning slik at det skal være enklere for barna å forstå. Testleder innleder med å si at hun skal lese en setning og at barnet skal hjelpe henne med å finne det motsatte ordet av det hun sier. Prøveoppgaven er som følger: *Er kapteinen liten eller er han...?(stor)* (Ottem & Frost, 2005). I denne oppgaven er det ikke barnets minne som skal måles, derfor er det her tillatt med gjentakelser. De tre første oppgavene er av samme type som prøveoppgaven. Etter det kommer oppgaver hvor testleder sier; hva

er det motsatte av... (et bestemt begrep). Et eksempel er: *Hva er det motsatte av starte?* (Ottem & Frost, 2005).

Også her skal testleder avbryte etter tre påfølgende feil. Det gis 1 poeng for riktig svar og 0 poeng for galt svar (Ottem & Frost, 2005). Her kan det også være nødvendig med tilpasning av oppgavene til dialekter. Det er ikke alle steder i landet det er naturlig å omtale en jakke som *pen*, men det ville kanskje være mer naturlig å si: *jakken er fin*. Fin og pen er begge antonymer til stygg (*Er jakken pen/fin eller er den ... (stygg)?*). Maksimal poengsum på denne oppgaven er 14. Denne poengsummen skal slås sammen med skåren på *ordforklaring* som er den andre oppgaven i deltesten *begreper*.

Ordforklaring

Denne oppgaven skal også måle barnas *begrepsmessige utvikling* (Ottem & Frost, 2005). Barnet skal gi en definisjon på talte ord, dette måler barnets ”språklige kunnskaper ervervet over tid” (Ottem & Frost, 2005, p. 8). Testleder forteller barnet at hun vil stille ham noen spørsmål. Prøveoppgaven blir presentert: *Hva er en hatt?* (Ottem & Frost, 2005). Her er det lagt til tre oppgaver i begynnelsen av oppgaven: *Hva er en katt, hva er en gave, og hva er en telefon?* Barna får 0 poeng for utilstrekkelig forklaring, feil svar eller ingen svar, og 1 poeng for riktig svar. For at svaret skal være riktig, må det være en betydning av ordet som man kan finne i vanlige ordbøker (Ottem & Frost, 2005). Oppgaven avbrytes etter tre 0 poengssvar etter hverandre. Maksimal poengsum er her 15. Sammenlagt med *motsetninger* er maksimal poengsum for oppgaven *begreper* 29 poeng.

Når man har med begrepstester å gjøre er det alltid en viss form for skjønn involvert i skåringen. Testleder vil få et inntrykk av om begrepet er forstått eller er i nærheten av å være forstått, og kan da gi poeng. Det kan være vanskelig å gjengi denne forståelsen til andre. Selv om man skulle ha et lydopptak av situasjonen, er det ikke alt et lydopptak kan formidle. Nonverbalt språk spiller også en stor rolle i enhver kommunikasjonssituasjon, som en testsituasjon også kan sies å være.

3.6 Skjemaet “20 spørsmål om språkferdigheter”

”20 spørsmål om språkferdigheter” er en liste som inneholder 20 uttalelser om barnets språklige fungering, som førskolelæreren skal rangere fra 1 til 5 etter hvor godt de passer til det enkelte barn. 1 betyr helt galt, og 5 betyr helt riktig. Det vil si at jo høyere skåre barnet får på dette skjemaet, jo større problemer har de i følge førskolelæreren. Påstandene er delt inn i tre problemområder; *semantisk evne*, *vansker med å motta et budskap* og *vansker med å formidle et budskap*. Vansker med å motta/formidle et budskap handler om barnets kommunikative evner (Ottem, 2009). Skjemaet inneholder også 7 uttalelser om barnets relasjoner til andre, dette er tatt med for å kunne få et mer helhetlig bilde av barnet, samtidig som en del barn som strever med språket kan ha vansker med å forholde seg til andre (Botting & Conti-Ramsden, 2000). Skjemaet er laget for barn mellom 5 og 16 år, dette er et stort aldersspenn og barna går gjennom en enorm utvikling på denne tiden, men førskolelæreren/læreren skal vurdere barnet opp mot andre barn på samme alder. Og som nevnt i kapittel 2.6.1 så har analyser vist at skjemaet er aldersuavhengig.

”20 spørsmål om språkferdigheter” er ment å være et supplement til tester brukt til utredning av språkvansker hos barn og unge (Ottem, 2009). Ottem (2009) fant at det ikke var signifikante korrelasjoner mellom delskalaene ”vansker med å motta et budskap”, ”vansker med å formidle et budskap” og screeningtesten Språk 6-16. Dette mener han er en indikasjon på at skjemaet gir informasjon om kommunikative ferdigheter, som ikke fanges opp av testen.

I skjemaet blir førskolelærerne først bedt om å gi litt generell informasjon om barnet, som kjønn, eventuelle vansker, hvor barnet er født og foreldrenes morsmål, barnets morsmål og barnets hjemmespråk (se vedlegg nr 1). Deretter kommer de 20 spørsmålene, eller utsagnene, om språkferdigheter. Respondentene instrueres til å vurdere hvor godt disse utsagnene passer til det aktuelle barnet sett i forhold til andre barn på samme alder. Førskolelærerne blir oppfordret til å bruke hele skalaen fra 1 til 5 og å huske å svare på alle utsagnene, selv om de ikke er hundre prosent sikre eller synes noe virker rart. Det gjøres for å forsøke å få svar på alle spørsmålene.

Manglende svar svekker spørreskjemaets styrke og generaliserbarhet (Christophersen, 2006; De Vaus, 2002). Utsagnene er delt inn i semantisk evne (utsagn 1-8), vansker med å motta et budskap (utsagn 9-14) og vansker med å formidle et budskap (utsagn 15-20). Deretter kommer 7 utsagn om barnets relasjoner til andre. Også her blir førskolelærerne instruert til å bruke hele skalaen, svare på alle utsagnene og å se det i forhold til andre barn på samme alder.

Spørreskjemaet er selvadministrerende, og førskolelærerne styrte selv når de ville svare på dem. Skåringen av skjemaene er dermed ikke gjort i samarbeid med testleder, men noen av førskolelærerne har samarbeidet om å svare på dem. Hvis førskolelærerne hadde noen spørsmål var studentene tilgjengelig for dem. Det er ikke mange spørsmål vi har fått, noe som tyder på at skjemaet er ganske selvforklarende. Det foreligger ikke normer for hvilke skårer som tyder på vansker eller ikke vansker hos barn i den aldersgruppen som er med i denne studien, men ettersom vurderingen av barna ved hjelp av dette skjemaet skal være aldersuavhengig kan en ta utgangspunkt i den ”cutoff”- skåren, eller kritiske grensen som Ottem (2009) fant i sin studie. Ottem hadde data fra typiske barn, og barn som hadde påviste språkvansker. Ved å sammenligne skårene til disse to gruppene fant han ut at en kritisk grense på 31 poeng (31%) var hensiktsmessig og at ca 83% av barna som hadde påviste språkvansker befant seg over denne grensen (Ottem, 2009).

3.7 Analyse

Data skal analyseres ved hjelp av statistikkprogrammet SPSS 16.0 (Statistical Package for the Social Sciences). For å analysere data vil det bli benyttet både deskriptiv og analytisk statistikk. Sammenhengen mellom de to variablene Språk 5-6 og ”20 spørsmål om språkferdigheter” vil bli undersøkt ved en korrelasjonsanalyse og gjennom bruk av en krystabell. For å gjøre krystabellen mer oversiktlig har data blitt gruppert i femtrinnskalaer. Én skala for sumskårene på screeningtesten Språk 5-6 og én for sumskårene på ”20 spørsmål om språkferdigheter”. Skårene fra Språk 5-6 er

gruppert etter de kategoriene som oppgis i manualen til Språk 6-16 (Ottem & Frost, 2005, p. 10):

- 1) Risikogruppe for språkvansker (-1,5 standardavvik under gjennomsnittet og dårligere)
- 2) Under gjennomsnittet (-1,5 til -1 standardavvik under gjennomsnittet).
- 3) Nedre normalområde (-1 til 0 standardavvik under gjennomsnittet).
- 4) Øvre normalområde (0 til +1 standardavvik over gjennomsnittet).
- 5) Over normalområdet (+1 standardavvik over gjennomsnittet og bedre).

Sumskårene fra ”20 spørsmål om språkferdigheter” er gruppert etter den inndelingen som Ottem (2009) gjorde da han validerte skjemaet. Disse er gruppert etter procenter av maksimal skåre. Her er det viktig å huske på at maksimal skåre er 100, men minimumsskåren er 20 og ikke 0.

- 1) 20 % av maksimal skåre.
- 2) 21 - 30 % av maksimal skåre.
- 3) 31 – 40 % av maksimal skåre.
- 4) 41 – 50 % av maksimal skåre.
- 5) 51 – 100 % av maksimal skåre.

(Ottem, 2009, p. 17).

Sensitivitet og spesifisitet vil bli beregnet slik som Rydz et al. (2005) beskriver det i sin artikkel og som nevnt i kapittel 2.7. Sensitivitet = Sann positiv/ (sann positiv-falsk negativ) og spesifisitet = sann negativ /(sann negativ + falsk positiv) (Rydz, et al., 2005, p. 7).

$$\text{Sensitivity} = \frac{\text{True-positive}}{\text{True-positive} + \text{false-negative}}$$

$$\text{Specificity} = \frac{\text{True-negative}}{\text{True-negative} + \text{false-positive}}$$

Krysstabellen vil danne grunnlag for klassifiseringen. Falsk positiv vil være gruppe nr. 1 og 2 på testen og gruppe nr. 1 og 2 på spørreskjemaet. Sann positiv vil være gruppe nr. 1 og 2 på testen og gruppe nr. 3, 4 og 5 på spørreskjemaet. Falsk negativ vil være gruppe nr. 3, 4, og 5 på testen og gruppe nr. 3, 4 og 5 på spørreskjemaet. Sann negativ vil være gruppe nr. 3, 4, og 5 på testen og gruppe nr 1 og 2 på spørreskjemaet. Resultatene fra spørreskjemaet vil regnes for den ”sanne” klassifisering, og resultatene fra Språk 5-6 vil bli målt opp mot den ”sanne” klassifisering for å se hvor nøyaktig testen er. Vurderingen av resultatene fra spørreskjemaet som ”sanne” bygger på at det er spørreskjemaet som brukes som kriterium for å vurdere testen. Selv om Ottem og Frosts (2005) inndeling setter bare de i den nederste gruppen ($\leq -1,5$ SD) i en risikogruppe for språkvansker, så befinner cutoffskåre for å si at noen har *spesifikke språkvansker* seg i den nest nederste gruppen (-1 til $-1,5$ SD), da den ligger på $-1,25$ standardavvik under gjennomsnittet (Leonard, 1998). Bishop (1997) mener vansken bør være så alvorlig at barnet skårer minst ett standardavvik under gjennomsnittet på aldersnormerte prøver for å kunne si at det handler om en spesifikk språkvanke. Derfor vil det være de to nederste gruppene *under gjennomsnittet* (-1 til $-1,5$ SD) og *risikogruppe for språkvansker* ($\leq -1,5$ SD) som blir tatt for å være resultater på testen som tyder på at barna er positiv for vansker.

3.8 Validitet og reliabilitet i undersøkelsen

Denne oppgaven omhandler validitet. I dette kapittelet er det testresultatenes gyldighet som skal vurderes, og ikke testens gyldighet. Validitet handler om hvor gyldige resultatene av målingen kan sies å være (Befring, 2007). Altså om prosessen med å innhente informasjonen om barna har vært gjort slik at resultatene er lite

påvirket av systematiske målingsfeil (Kleven, 2002b). Validitet er en kontinuerlig prosess og kan dermed ikke bli fullstendig oppfylt (AERA, et al., 1999).

På grunn av at screeningtesten Språk 5- 6 er helt ny, hadde ingen av studentene kjennskap til testen fra før. Det ble derfor besluttet at studentene skulle gjennomføre noen prøvetestinger før de dro ut i barnehagene, på den måten ønsket man å unngå for mange systematiske målefeil, og dermed styrke testresultatenes validitet.

Arbeidet med å fylle ut spørreskjemaene ble forsøkt fordelt mellom førskolelærere i barnehagene, slik at det ikke skulle bli for mye på hver enkelt. Dette ble gjort av hensyn til at mengden arbeid ikke skulle bli for stor for hver enkelt, men også for å prøve å oppnå nøyaktighet i utfyllingen av skjemaene. Spørreskjemaer burde ikke være for lange fordi svært lange spørreskjemaer legger en stor byrde på respondentene og det kan føre til at de vegrer seg for å delta (De Vaus, 2002). Dette kan også relateres til at førskolelærerne som deltok i undersøkelsen ikke burde få for mange skjemaer å svare på hver. Dette var tross alt arbeid som de tok på seg utenom sin travle arbeidshverdag. Det inngikk i avtalen med barnehagene at vi skulle komme dit å teste og at førskolelærerne skulle svare på spørreskjemaene. Likevel ville nok et for langt spørreskjema kunne gjøre at noen barnehager ville vegre seg fra å delta i undersøkelsen. Skjemaet som en har benyttet i dette prosjektet vurderes ikke som å være for langt, men det har blitt tatt med i betraktning at det var mange skjemaer som skulle fylles ut, og en har derfor forsøkt å fordele mengden mellom førskolelærerne i barnehagene.

Vi ønsket å teste alle barna i barnehagen i den gitte aldersgruppen for å finne normalvariasjon, slik forsøkte en å unngå seleksjon, noe som kunne ha resultert i et svært skjevt utvalg.

3.9 Metodekritikk

Alle studier har begrensninger. Det kan for eksempel være vanskelig å få tak i personer som er villige til å delta i undersøkelsen, eller det kan være vanskelig å finne

personer som passer de utvalgskriteriene som er satt, spesielt hvis utvalgskriteriene er svært strenge. En annen begrensning som kan være aktuell er begrensede ressurser. Det kan være mangel på ressurser i form av penger, tid eller arbeidskraft. Derfor vil det være noe som kunne vært gjort annerledes for å oppnå et bedre resultat. Refleksjoner rundt hva som kunne vært gjort annerledes vil kunne være til nytte for andre som skal vurdere den aktuelle forskningen eller forsøke å gjenskape undersøkelsen.

I denne undersøkelsen kunne tidspunktet brevene til foreldrene ble sendt ut på vært valgt med mer omhu. Disse ble stort sett utlevert rett før jul, som er en travel tid for alle, også barnehagene. Kanskje var det slik at noen ikke ville delta fordi de syntes det ble for mye å tenke på i en ellers stresset hverdag. Dette fikk ikke store følger for vårt prosjekt, da vi fikk tak i nok testpersoner, men spørsmålet er om vi har fått tak i et representativt utvalg. Vi hadde som mål å teste ca 25 barn hver, noe som skulle være rikelig ettersom vi regnet med at noen av ulike årsaker måtte filtreres ut og vi gjerne ville lande på et utvalg som vi kunne bruke med ca 110 barn. Hvis vi hadde testet 25 hver ville vi hatt et utvalg på 125 stk. Vi fikk et utvalg på 118 stykker, dette grunnet at noen barn var borte på grunn av sykdom eller ferie når testingen fant sted.

Det hadde vært ønskelig å kunne ha testet en gruppe med barn som hadde påviste språkvansker. På den måten kunne en lettere ha sagt noe om testens evne til å klassifisere barna som *med* eller *uten* språkvansker. Da kunne en kanskje også ha sagt noe om hvordan sensitiviteten og spesifisiteten ville vært om man kombinerte de to instrumentene. Det ville også ha vært nyttig å ha et mål på barnas intelligens, slik at man hadde hatt bedre grunnlag for å vurdere om barna har spesifikke språkvansker eller ikke.

3.10 Etikk

De etiske hensyn som må tas i en forskningsundersøkelse varierer fra prosjekt til prosjekt. I dette prosjektet har informantene krav på at opplysningene som samles inn om dem blir oppbevart på en forsvarlig måte, slik at ikke uvedkommende kan få tak i

dem (Befring, 2007; NESH, 2006). Opplysningene skal også anonymiseres slik at ingen skal kunne vite hvem det er opplysninger om (Befring, 2007; NESH, 2006) . Når prosjektet avsluttes skal alle protokoller og andre opplysninger makuleres, og lydopptakene slettes (Befring, 2007; NESH, 2006).

Ettersom det er barn som har vært informanter i dette prosjektet krevdes samtykke fra foreldrene til at barna deltok i prosjektet. Foreldrene må få informasjon som er tilstrekkelig og forståelig for dem, slik at de kan gi et informert samtykke til at deres barn kan delta i studien (Befring, 2007; NESH, 2006). De må også forsikres om at det ikke vil medføre noen følger for dem eller barna hvis de velger å ikke delta i studien, eller velger å trekke seg underveis ((Befring, 2007; NESH, 2006).

Selv om foreldrene har gitt sitt samtykke til at barna kan være med, har vi vurdert det som viktig at barna selv også ønsker å være med. Derfor har vi vært forberedt på å avbryte testing eller unnlate å teste barn som vegrer seg fra å delta (Befring, 2007). Dette ble derimot ikke nødvendig da de fleste barna gledet seg til å få være med.

Listene med navn på barna som er testet blir fortsatt oppbevart i barnehagene, disse er bedt om å beholde disse til prosjektet er ferdig (Befring, 2007). Dette er gjort fordi vi ønsker å kunne gi tilbakemelding til barnehagene hvis vi finner noen barn som viser tegn til språkvansker gjennom de instrumentene vi har benyttet.

4. Resultater fra undersøkelsen

Denne undersøkelsen har som mål å validere Språk 5-6. Det vil bli gjort ved å se på om den nye testen holder mål mot spørreskjemaet. Dette gjøres ved å se på sammenhengen mellom de to kartleggingsinstrumentene. Men før dette kan gjøres må det undersøkes om måleinstrumentene har en indre konsistens. Det gjør man ved å regne ut instrumentenes reliabilitetskoeffisienter (Cronbachs Alpha). Utgangspunktet vil være en presentasjon av alle resultatene fra de to måleinstrumentene.

Sammenhengen mellom de to instrumentene vil undersøkes gjennom bruk av en korrelasjonsanalyse og en krystabell. En signifikant sammenheng mellom de to instrumentene vil bli tatt til inntekt for at testen viser tegn til å være mer økologisk valid enn antatt. Testen kan ses som mer økologisk valid i den forstand at den korrelerer med et instrument som måler barnas språkferdigheter på en mer økologisk valid måte. Ved å ta utgangspunkt i barnas spontane tale i hverdagssituasjonen kan spørreskjemaet betraktes som lite smittet av ”testsituasjonen”, og dermed mer økologisk valid. Til slutt vil det bli forsøkt å regne ut testens nøyaktighet til å kategorisere noen barn som positiv (med språkvansker) og negativ (uten språkvansker). Dersom det viser seg at testen er nøyaktig, vil det være med på å danne grunnlag for å stole på at de slutningene som blir gjort på bakgrunn av resultatene er gyldige, eller valide. I kapittel 5 blir det reflektert rundt disse resultatene og de vil bli sett i sammenheng med aktuell teori og empiri.

4.1 Deskriptiv statistikk

4.1.1 Resultater på screeningtesten og spørreskjema

Dette delprosjektets bidrag til hovedprosjektet er å se på sammenhengen mellom spørreskjemaet og den nye screeningtesten og med dette si noe om testens gyldighet. Som en del av vurderingen av testens validitet kan det å se på testens egnethet for aldersgruppen være nyttig, da det er første gang instrumentene benyttes på barn i denne aldersgruppen. Her vil resultatene som barna fikk på henholdsvis

screeningtesten Språk 5-6 og spørreskjemaet ”20 spørsmål om språkferdigheter” presenteres for å kunne gi grunnlag for drøfting av utvalget og for å kunne si noe om resultater i forhold til en oppdeling i aldersgrupper.

Tabell 3: Resultater på screeningtesten og spørreskjema

		N ⁵	M	SD	Min	Max	Skewness	Kurtosis
Hele utvalget	Språk 5-6	118	23,7	5,34	11	38	.180	-.170
	20 spørsmål	118	24,9	8,12	20	56	2.077	3.719
Barn 5,0 til 5,5 år:	Språk 5-6	66	22,9	5,23	11	38	.156	.210
	20 spørsmål	66	26	8,8	20	56	1.779	2.430
Barn 5,5 til 6,0 år:	Språk 5-6	52	24,7	5,35	14	37	.205	-.601
	20 spørsmål	52	23,5	6,98	20	53	2.703	7.442

6

Tabell 3 viser resultatene for hele utvalget samt for utvalget delt i to aldersgrupper. Tabellen inneholder informasjon om antall barn i gruppene (N), gjennomsnitt (M), standardavvik (SD), minste skåre (Min), maksimumsskåre (Max), fordelings skjevhet (skewness) og om fordelingen er spiss eller flat (kurtosis).

Skewness, eller skjevhet, sier noe om fordelings symmetri. Hvis fordelingen har en positiv skjevhet er den høyreskjev og hvis den har en negativ skjevhet, er den venstreskjev (Christophersen, 2006; Tabachnick & Fidell, 2001). Verdier for skewness som er fra -1 til 1 indikerer at fordelingen er lite skjev, fordelingen er moderat skjev hvis den har verdier for skewness inntil -2 og 2, fordelingen er sterkt skjev for verdier inntil -3 og 3 og svært sterkt skjev for verdier som ligger utenfor -3 og 3 (Christophersen, 2006, p. 47). Kurtosis sier noe om fordelingen er spiss eller flat. Hvis kurtosis for fordelingen er positiv er den spiss og hvis kurtosis er negativ er fordelingen mer flat (Christophersen, 2006; Tabachnick & Fidell, 2001). Det

⁵ N er hele utvalget. De to undergruppene som er laget på bakgrunn av alder vil bli betegnet med n i teksten.

⁶ Skewness = skjevhet

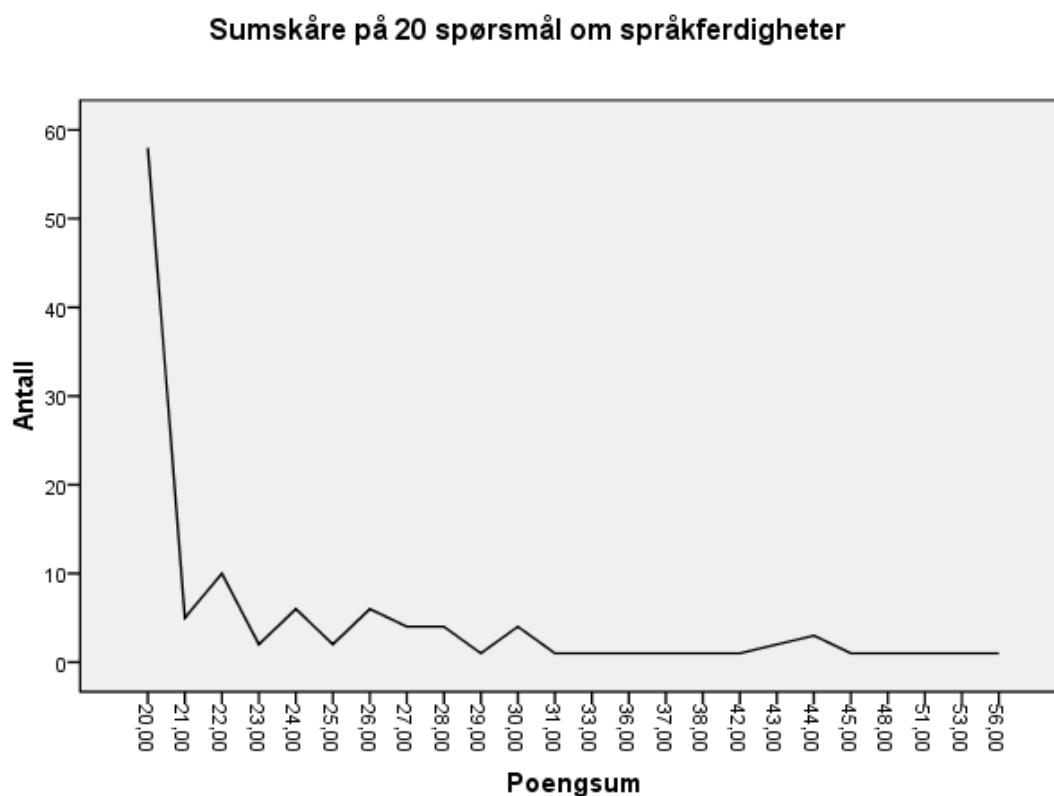
optimale er at fordelingen er normalfordelt. Verdier for kurtosis som befinner seg mellom -1 og 1 indikerer at fordelingen avviker lite fra normalfordelingen, det viser moderat avvik fra normalfordelingen med verder inntil -2 og 2, verdier inntil -3 og 3 viser sterkt avvik fra normalfordelingen og avvik utenfor -3 og 3 indikerer svært sterkt avvik fra normalfordelingen (Christophersen, 2006, p. 47).

Hele utvalget

Tabell 3 viser fordelingen på de to variablene for hele utvalget (N). Det er 118 barn med i utvalget. Barna i utvalget skårer i gjennomsnitt 23,7 poeng på screeningtesten Språk 5-6, og har standardavvik fra dette gjennomsnittet på 5,34. Skårene varierer fra 11 poeng, som er den laveste skåren, til 38 som er den høyeste skåren. Fordelingen for skårene på Språk 5-6 har en skjevhet som ligger nært 0 som vil si at den er tilnærmet normalfordelt. Skewness er her .180 som da antyder en svært liten skjevhet mot høyre og kurtosis indikerer lite avvik fra normalfordelingen. Her er kurtosis for skårene på Språk 5-6 til hele utvalget -.170, som indikerer at fordelingen er nesten normalfordelt, men at den er en anelse flat i henhold til angivelsene gitt av Christophersen (2006).

På ”20 spørsmål om språkferdigheter”-skjemaet er en skåre på 20 det minste man kan oppnå, og derfor er det den laveste skåren. Gjennomsnittet for utvalget er en skåre på 24,9 poeng, med standardavvik på 8,12. Den høyeste skåren som noen oppnådde var 56, som er i den høyeste problemkategorien og indikerer alvorlige vansker. Denne fordelingen har en svært lang hale. Fordelingens skjevhet er moderat. Kurtosis ligger derimot over grensen for svært sterkt avvik fra normalfordelingen. Om man ser på en grafisk fremstilling av denne fordelingen så ser man at den er svært spiss (figur 2).

Figur 2: Grafisk fremstilling av fordelingen for hele utvalget på spørreskjemaet ”20 spørsmål om språkferdigheter”.



På grunn av at prosjektets utvalgskriterier fordret et utvalg uten kjente språk-, hørsels- eller andre vansker som kan være årsak til språkvansker, fikk et stort flertall av barna (86,4 %) skårer som befant seg innenfor normalområdet. Det er noe av grunnen til at fordelingen ikke er normalfordelt. For å se på om fordelingen avhenger av barnets alder, ble utvalget delt i to aldersgrupper. En med barn fra 5,0 år til 5,5 år, og en med barn fra 5,5 år til 6,0 år.

Den yngste aldersgruppen

Den yngste aldersgruppen er den største (N= 66). Det vil si at 55,9 % av barna i det totale utvalget er under 5,5 år. Deres gjennomsnittsskåre på Språk 5-6 er 22,9 poeng, med standardavvik på 5,23. Gjennomsnittet til denne aldersgruppen er dermed litt lavere enn gjennomsnittet til hele gruppen. Den laveste skåren er 11, som også var den laveste skåren i hele utvalget. Den høyeste skåren er 38, som var den høyeste

skåren i hele utvalget. Man finner altså her en stor variasjon mellom skårene, og en finner både den høyeste skåren for hele utvalget og den laveste skåren for hele utvalget i den yngste aldersgruppen. Fordelingens skewness er på .156 som vil si at den er tilnærmet normalfordelt. Den er litt høyreskjev, men innenfor det Christophersen (2006) vil si er en *liten* skjevhet mot høyre. Kurtosis er også liten (.210) som også er et tegn på at fordelingen er tilnærmet normalfordelt, men en anelse spiss.

For skårene på ”20 spørsmål om språkferdigheter” er gjennomsnittet for den yngste aldersgruppen 26 poeng, med standardavvik på 8,8. Gjennomsnittet er høyere enn det for det samlede utvalget, men med et forholdsvis likt standardavvik. Den laveste skåren for denne aldersgruppen er 20 poeng, som er den laveste mulige skåren. Den høyeste skåren blant barna i den yngste aldersgruppen er 56 som også er den høyeste skåren i hele utvalget (N). Den store variasjonen i skårene kan tyde på at det er stor forskjell mellom barna i denne alderen, at de er i stor utvikling og derfor blir det store ulikheter. Fordelingen har en moderat skjevhet. Kurtosis er opp mot 3, noe som indikerer et sterkt avvik fra normalfordelingen, det positive fortegnet betyr at fordelingen er spiss (Christophersen, 2006). Det at fordelingen er spiss kan forklares ut fra at det er mange ”normaltfungerende” barn i utvalget og at disse følgelig vil skåre lavt på skjemaet og dermed skape en opphopning i den nedre del av skjemaet.

Den eldste aldersgruppen

Den andre undergruppen representerer barn fra 5,5 år til 6,0 år. 44,1 % av hele utvalget av barn befinner seg i denne gruppen (N = 52). For skårene på screeningtesten Språk 5-6 er gjennomsnittet 24,7 poeng, det er høyere enn gruppen som samlet og undergruppen med yngre barn. Gjennomsnittet til den eldste gruppen ligger 1 poeng høyere enn gjennomsnittet for hele utvalget og 1,8 poeng høyere enn gjennomsnittet for den yngste gruppen.

Den laveste skåren på Språk 5-6 for denne gruppen er på 14 og det er 3 poeng høyere enn den laveste for hele utvalget. Den høyeste skåren er på 37 og det er 1 poeng lavere enn den høyeste for hele utvalget. Skewness på .205 viser at også denne

fordelingen er tilnærmet normalfordelt, men med en liten skjevhet mot høyre. Som vil si at det er en opphopning til venstre for gjennomsnittet og at fordelingen har en lengre hale mot høyre. Kurtosis for denne fordelingen er $-.601$ som vil si at den er lite avvikende fra normalfordelingen, men litt flat.

For skårene på spørreskjemaet er gjennomsnittet 23,5 poeng som er lavere enn både gruppen som helhet og den yngste gruppen. Standardavviket er på 6,98 som er en del lavere enn det er i hele utvalget og i den yngste gruppen. Det viser at variasjonen i den eldste gruppen er mindre enn i den yngste gruppen og i utvalget som helhet, det vil si at de eldste barna skårer jevnere enn de yngste barna. Den laveste poengsummen er 20, som er det samme som for hele utvalget og de yngste barna. . Den høyeste skåren i denne gruppen var 53 som er lavere enn den høyeste skåren for hele utvalget, men som fremdeles er over grensen for å vise tegn til store språkvansker, som ligger på 51 poeng (Ottem, 2009).

Skewness for denne fordelingen er på 2.703 som kan sies å være en sterkt skjev fordeling, og som nevnt tidligere skyldes dette at det er mange som får minimumsskåren. Kurtosis ligger på 7.442 som indikerer et svært sterkt avvik fra normalfordelingen, den er svært spiss. Skjemaet er konstruert for å finne barn med vansker og på grunn av at barna i utvalget for det meste er vanlige barn uten vansker vil disse klumpe seg nederst på skalaen. Det ser også ut til at de eldste barna er særlig flinke og har få språklige vansker. Derfor er fordelingen for de eldste barna ekstra spiss.

En ser her at fordelingen til screeningtesten Språk 5-6 er tilnærmet normalfordelt for alle gruppene, men at fordelingen til spørreskjemaet ”20 spørsmål om språkferdigheter” ikke er det. Reliabiliteten må undersøkes for å se om det kan rettferdiggjøres å bruke spørreskjemaet som kriterium for vurdering av testens validitet. Som en del av valideringen av testen skal dennes reliabilitet også undersøkes. Reliabiliteten undersøkes ved å se på indre konsistens i testen og spørreskjemaet, dette blir gjort ved hjelp av Cronbachs Alpha (Gall, et al., 2007). Denne måten å vurdere reliabilitet på, er valgt på grunn av at det kun er tatt én måling

med hvert instrument. En annen måte å vurdere reliabilitet på ville vært å se på konsistensen i samme persons utfall på henholdsvis testen og spørreskjemaet hvis den samme personen var målt på disse to ganger.

Ettersom det i dette prosjektet er benyttet data som er samlet inn av flere personer kunne en eventuelt ha beregnet vurdererreliabilitet, men det forutsetter at de samme testene er skåret av flere personer for så å sammenligne resultatene (Kleven, 2002b). Det var det ikke rom for i dette prosjektets omfang, slik at dette ikke har vært en mulighet.

4.1.2 Reliabilitetsanalyser

Som nevnt er det viktig å undersøke testens reliabilitet for å kunne vite noe om hvor pålitelige resultatene en får fra den er. Spørreskjemaets reliabilitet må undersøkes for å se om det kan rettferdiggjøres å bruke det som kriterium å måle testen opp mot. Her vil det først bli redegjort for noen forutsetninger for analysen, deretter presenteres en tabell med oversikt over reliabilitetskoeffisientene til hver enkelt deltest i Språk 5-6 og totalen, samt reliabilitetskoeffisienten til ”20 spørsmål om språkferdigheter” og tilleggsskalaen ”relasjoner til andre”. I denne tabellen er reliabilitetskoeffisientene presentert både for hele utvalget og for utvalget delt i to aldersgrupper.

Forutsetningen for at man kan foreta en reliabilitetsanalyse av indre konsistens i screeningtesten, er at det finnes varians i resultatene til de enkelte items (Crocker & Algina, 1986). Det viste seg da at noen items fra testen måtte utelukkes da de ikke viste varians. De items som ikke hadde varians var de enkleste oppgavene og de vanskeligste. Det betyr at alle barna har klart de enkleste oppgavene, og som forventet har ikke barna kommet så langt at de har svart på de vanskeligste oppgavene, da disse er beregnet for barn som er mye eldre enn dem. Oppgavene er beregnet på barn opptil 16 år.

For deltesten setningsminne klarte alle barna de tre første oppgavene, altså tre av fire nye oppgaver som er lagt til for å få med seg de yngste barna. De tre første oppgavene er: *Gutten løper*, *Mannen spiser*, *Jenta er glad*. I denne deltesten var det

de seks vanskeligste items som ingen klarte å svare på. Til sammen var det 9 items som ble utelukket fra analysen fra denne deltesten.

For deltesten ordspenn var det bare de vanskeligste oppgavene som ikke hadde varians, det vil si at de enkleste oppgavene var tilstrekkelig vanskelig til at enkelte barn svarte feil på dem, i motsetning til setningsminne. . For denne deltesten var det de tre vanskeligste oppgavene barna i dette utvalget ikke klarte, og som dermed ikke ble tatt med i analysen.

Deltesten begreper er som nevnt delt i to oppgaver: Motsetninger og ordforklaring. Oppgaven motsetninger var vanskelig for barna mellom 5 og 6 år. Her var det varians i alle de 5 første oppgavene, men item nr. 6 (*Hva er det motsatte av tam?*) var det ingen som klarte, heller ikke item nr. 9 (*Hva er det motsatte av sende?*). I tillegg var det ingen som klarte de 4 vanskeligste items. Disse fikk dermed ingen varians og ble derfor utelukket fra analysene. For oppgaven ordforklaring var det de fire vanskeligste oppgavene som ikke hadde varians og ble utelukket fra analysene. Fra deltesten begreper ble det til sammen utelukket 10 items fra reliabilitetsanalysen. De items som ble utelukket fra reliabilitetsanalysen i de enkelte deltestene er som følger:

Tabell 4: Oversikt over items som ikke hadde varians

Setningsminne	Item nr. 1	Item nr. 2	Item nr. 3	Item nr. 15	Item nr. 16	Item nr. 17	Item nr. 18	Item nr. 19.	Item nr.20
Ordspenn	Item nr. 14	Item nr. 15	Item nr. 16						
Motsetninger	Item nr. 6	Item nr. 9	Item nr. 11	Item nr. 12	Item nr. 13	Item nr. 14			
Ordkunnskap	Item nr. 12	Item nr. 13	Item nr. 14	Item nr. 15					

Reliabilitetsanalyser basert på de items som hadde varians:

Tabell 5: Reliabilitet (Chronbachs alpha)

Språk 5-6	Setningsminne	Ordspenn	Begreper	Sum	20 spm	Relasjoner
5,0 – 5,5 år	.744	.722	.633	.825		
5,5 – 6,0 år	.799	.662	.637	.849		
Totalt	.762	.701	.642	.838	.910	.850
Antall items	11	13	19	43	20	7

En ser av tabell 5 at reliabiliteten for *setningsminne* er høyere for den eldste aldersgruppen enn den yngste, den er også høyere enn for de to gruppene sammenlagt. På *ordspenn* er det den yngste gruppen som har den beste reliabiliteten, også denne høyere enn reliabiliteten for hele utvalget. På deltesten *begreper* oppnår de to gruppene ganske like reliabilitetskoeffisienter, her er reliabiliteten til det samlede utvalget høyere enn de to gruppene hver for seg. På summen av testen har den yngste aldersgruppen lavest reliabilitet, den eldste aldersgruppen har høyere reliabilitet enn både den yngste gruppen og utvalget som helhet. Sumskårene av testen har tilstrekkelig reliabilitet uavhengig av om de deles i to aldersgrupper eller ses under ett. Det vil si at de alle befinner seg over .8 og kan dermed benyttes i denne sammenhengen (Gall, et al., 1996; Lichtenstein & Ireton, 1984).

Spørreskjemaet ”20 spørsmål om språkferdigheter ” har en svært høy reliabilitetskoeffisient (.910) som kan sies å være godt nok til at instrumentet kan brukes som kriterium for valideringen av screeningtesten Språk 5-6. Også delskalaen ”relasjoner til andre” har en god reliabilitet (.850) som gjør at også denne kan sies å være tilstrekkelig reliabel til å benyttes i denne sammenhengen.

⁷ Reliabiliteten til de enkelte deltestene i Språk 5-6 er tatt med her for å kunne sammenligne dem med tall fra Ottem og Frost(2005).

Konklusjonen blir da at instrumentenes reliabilitet er tilstrekkelig til at det er forsvarlig å bruke dem i denne sammenhengen. Samtidig er det viktig å huske på at screeningtesten fremdeles er under utvikling og at det er sannsynlig at den vil justeres på bakgrunn av funn som blir gjort i forbindelse med utprøvingen av den, som dette prosjektet er en del av. Hvis det blir foretatt endringer må det undersøkes reliabilitet på nytt.

4.2 Analytisk statistikk

4.2.1 Analyse av forholdet mellom screeningtesten Språk 5-6 og spørreskjemaet "20 spørsmål om språkferdigheter".

Denne oppgavens problemstilling er: *Hvilket forhold er det mellom førskolelæreres vurdering av barns språklige ferdigheter og disse barnas resultater på Språk 5-6?*

Denne problemstillingen er valgt for å kunne belyse testens validitet. Første ledd i validering av screeningtesten Språk 5-6 handler om hvorvidt testen måler det den skal måle. Altså skal vi forsøke å finne ut om testen måler barnas virkelige språklige fungering slik den ville fremtre i en naturlig situasjon. Dersom det finnes en signifikant sammenheng mellom de to instrumentene tas det som et argument for at testen fanger barnas virkelige språklige ferdigheter. Forholdet mellom de to instrumentene vil først bli undersøkt ved hjelp av en korrelasjonsanalyse. Ettersom fordelingen for spørreskjemaet "20 spørsmål om språkferdigheter" er svært spiss vil det ikke være gunstig å bruke Pearsons R som korrelasjonskoeffisient. Derfor benyttes Spearmans Rho i stedet. Forholdet mellom de to instrumentene vil deretter analyseres i en krystabell hvor resultatene er delt inn i to femtrinns skalaer. Det vil også her bli undersøkt om det er en signifikant sammenheng mellom de to instrumentene ved bruk av χ^2 , styrken på en eventuell sammenheng vil bli beregnet ved Cramers V.

4.2.2 Korrelasjonsanalyse

Tabell 6: Korrelasjonsmatrise

	20 spørsmål om språkferdigheter
Språk 5-6:	
5,0 – 5,5 år	-.346**
5,5 – 6,0 år	-.365**
Totalt	-.374**
**Korrelasjonen er signifikant på 0.01 nivå	

Det betyr at det er opptil 1% sjanse for at den sammenhengen som er funnet er en ren tilfeldighet. Dette vil si at det er minst 99% sjanse for at dette er et reelt funn.

Korrelasjonsmatrisen i tabell 6 viser at det er en negativ korrelasjon mellom de to instrumentene. Dette er naturlig da en lav skåre på spørreskjemaet indikerer at barnet ikke har vansker, mens en høy skåre på testen indikerer at barnet ikke har vansker, og omvendt. Korrelasjonen er signifikant på 0.01 nivå. Spearmans Rho for det samlede utvalget er -.374 som, dette er en moderat sterk korrelasjon (Connolly, 2007). For å se om det var noen forskjell mellom aldersgruppene ble analysen også foretatt på de to aldersgruppene separat. For barn på 5,0 til 5,5 år var da Spearmans Rho lik -.346, og for barn på 5,5 til 6,0 år var Spearmans Rho lik -.365. Begge var signifikant på 0.01 nivå. Det er en liten forskjell men den er ikke nevneverdig. Det er en signifikant korrelasjon mellom de to instrumentene og en kan derfor argumenter for testens validitet. Styrken på korrelasjonen er moderat, en kunne kanskje ønske at den var noe sterkere.

4.2.3 Krysstabell

For å gjøre krysstabellen mer oversiktlig har skårene blitt gruppert i femtrinnskalaer. Skalaen for skårene på testen er laget ut fra z-skårer og gruppert etter den skalaen som ble funnet hensiktsmessig og som ble brukt i manualen for screeningtesten Språk

6-16 (Ottem & Frost, 2005, p. 10). Skalaen som skårene fra spørreskjemaet ble gruppert etter er det samme prinsippet som Ottem fant hensiktsmessig da han validerte spørreskjemaet (Ottem, 2009, p. 17).

Tabell 7: Krysstabell Språk 5-6 og 20 spørsmål om språkferdigheter.

Språk 5-6		← -1.5		-1.5 til -1		-1 til 0		0 til +1		+1 →		Total	
20 spm:	20%	2	3,4%	3	5,2%	13	22,4%	26	44,8%	14	24,1%	58	100%
21-30 %		3	6,8%	3	6,8%	18	40,9%	18	40,9%	2	4,5%	44	100%
31-40%		0	0%	1	20,0%	3	60,0%	1	20,0%	0	0%	5	100%
41-50%		2	25,0%	3	37,5%	2	25,0%	1	12,5%	0	0%	8	100%
51 % →		1	33,3%	0	0%	2	66,7%	0	0%	0	0%	3	100%
Total		8	6,8%	10	8,5%	38	32,2%	46	39,0%	16	13,6%	118	100%

Tabell 7 viser sammenhengen mellom spørreskjemaet og testen. Det er en signifikant sammenheng mellom testen og skjemaet (Chi-square = 36,16, df = 16, p = .002).

Styrken på sammenhengen blir beregnet ved Cramers V som er lik .279, denne sammenhengen blir betegnet som *svak* (Connolly, 2007; Pallant, 2005).

Dersom en tar utgangspunkt i testen ser en at det er 8 stykker som har kommet i den laveste kategorien på testen, noe som indikerer at de ligger i en risikogruppe for språkvansker. Hele fem av de 8 får skårer på spørreskjemaet som ligger under den kritiske grensen (31%) for hva som indikerer vansker. Her er spørreskjemaet og testen helt uenig. Testresultatene viser at disse fem er i *risikozonen for språkvansker* og skårer betydelig lavere enn normalen for jevnaldrende barn, mens i følge spørreskjemaet har disse barna ingen vansker med språket. De tre siste barna har resultater på både testen og spørreskjemaet som setter dem i en vanskegruppe. På spørreskjemaet har disse barna oppnådd relativt høye skårer som indikerer en høy grad av vansker.

Det er litt flere som har skårer som plasserer dem i gruppen *under gjennomsnittet* på testen, men selv her er det 6 stykker som blir plassert i normalområdet av skårene på spørreskjemaet. Den ene av de fire som får problemskåre havner i gruppen rett over den kritiske grensen og de tre resterende i den nest høyeste vanskegruppen. Disse resultatene tyder på at testen og spørreskjemaet er relativt enige da testen egentlig har bare én vanskegruppe og spørreskjemaet har tre.

I nedre normalområdet for testskårene (-1 til 0) er de fleste barna kategorisert innenfor det som er det normale på spørreskjemaet, altså lavere skåre enn 31%. En ser derimot også at det er to stykker som skårer i den øverste vanskegruppen på skjemaet, som også befinner seg her.

I øvre normalområde for testskårene er det, som ventet, flest i de to laveste gruppene fra skjemaet, opptil 20 % og 21 til 30 %. Der er likevel to stykker som har blitt plassert i vanskegruppen av førskolelærerne (over 31 %).

I gruppen som befinner seg over normalområdet for testskårene har alle barna skårer under den kritiske grensen på 31 % i spørreskjemaet. Det viser at testen og førskolelærerne har konsensus om at disse barna ikke har noen særlige vansker med språket.

Sammenhengen mellom de to instrumentene er signifikant og en kan derfor argumentere for testens validitet. Styrken på denne sammenhengen er derimot ikke så sterk som en kunne ønske.

4.2.4 Analyser av testens nøyaktighet

Resultatene fra testen kan sies å være en vurdering. Ved å sette en kritisk grense eller cutoffskåre så vil de som havner under grensen på testen bli satt i gruppen som har vansker, mens de som er over den grensen vil bli satt i gruppen uten vansker. Disse plasseringene må kunne bli foretatt på det mest mulig nøyaktige grunnlaget for at en skal kunne stole på resultatene og at de slutningene som trekkes på bakgrunn av dem skal kunne være valide.

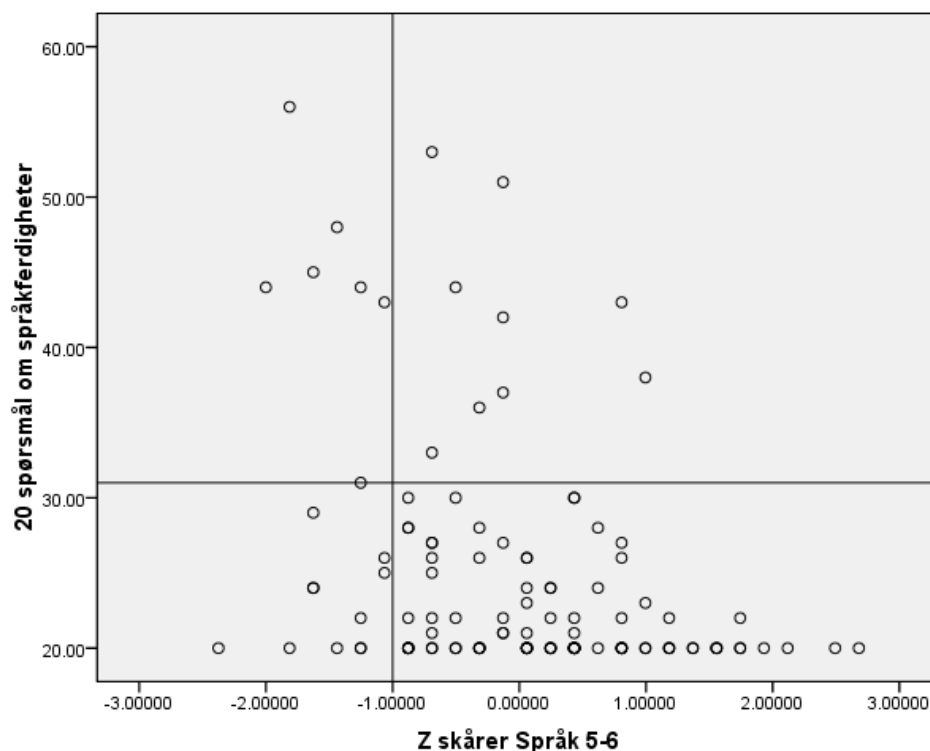
Data i denne undersøkelsen er ikke ideelle for denne typen analyse, da dataene er samlet inn samtidig og utvalget inneholder barn uten antatte vansker. Derfor vil resultater fra denne analysen bli tolket med forsiktighet. Det tas utgangspunkt i krysstabellen (tabell 4) for å plassere resultatene i kategorier. Kategorien positiv for vansker vil omfatte mer enn den som Ottem og Frost beskriver i manualen for Språk 6-16 (Ottem & Frost, 2005). Det er på grunn av at cutoffskåren, grensen for de som skårer $-1,25$ SD under M , tas med i gruppen *positiv for vansker*, da denne cutoffskåren er vanlig å operere med i forbindelse med diagnostisering av språkvansker. Dermed inneholder gruppen av de barna som tester positiv for vansker på Språk 5-6 de som har skårer som ligger allerede ved -1 SD og nedover.

Barna plasseres i 4 kategorier:

- 1) *Falsk positiv*: Denne kategorien innbefatter de som skåret fra -1 standardavvik under gjennomsnittet, på Språk 5-6, og nedover, og som skåret opp til 31% på 20 spørsmål om språkferdigheter skjemaet: 11 barn (9 %)
- 2) *Sann positiv*: Denne kategorien består av de som skåret fra -1 under gjennomsnittet og nedover på Språk 5-6, og skåret over 31% på 20 spørsmål om språkferdigheter skjemaet: 7 barn (6 %)
- 3) *Falsk negativ*: Denne kategorien inneholder de som skåret fra -1 standardavvik undergjennomsnittet og oppover på Språk 5-6, og som skåret over 31% på 20 spørsmål om språkferdigheter skjemaet: 9 barn (8 %)
- 4) *Sann negativ*: Denne kategorien omfatter de som skåret fra -1 standardavvik undergjennomsnittet og oppover på Språk 5-6, og som samtidig skåret mindre enn 31% på 20 spørsmål om språkferdigheter: 91 barn (77 %)

Dette kan illustreres ved et scatterplot hvor en har satt et skjeringspunkt på -1 SD på Språk 5-6 og et skjeringspunkt på 31 % på "20 spørsmål om språkferdigheter".

Figur 3: Grafisk fremstilling av utfallet fra testen og spørreskjemaet



Disse kategoriene er grunnlaget når testens sensitivitet og spesifisitet skal beregnes. Som vist i kapittel 3.7, er regnestykkene som følger:

$$\text{Sensitivitet} = \text{Sann positiv} / (\text{Sann positiv} + \text{falsk negativ}) = 7 / (7 + 9) = 7/16 = \underline{\underline{0,4375}}$$

$$\text{Spesifisitet} = \text{Sann negativ} / (\text{sann negativ} + \text{falsk positiv}) = 91 / (91 + 11) = 91/102 = \underline{\underline{0,892}}$$

Det vil si at testen har en sensitivitet på ca 43,8%. Testen finner altså ca 43,8% av de som har språkvansker, når kriteriet den blir målt opp mot er spørreskjemaet "20 spørsmål om språkferdigheter". Testens spesifisitet er på 89,2%. Det betyr at testen identifiserer 89,2% av de som ikke har språkvansker som negative, altså uten språkvansker. Om en legger sammen *sann positiv* og *sann negativ* finner en at testen plasserer 83% av barna i riktig "bås", mens om man legger sammen *falsk positiv* og *falsk negativ* finner en at testen plasserer 17% av barna i feil "bås".

5. Diskusjon

Resultatene fra undersøkelsen har nå blitt presentert. De viste at fordelingen for screeningtesten var tilnærmet normalfordelt. Det var noen forskjeller mellom de to aldersgruppene, men de var ikke store. Den eldste aldersgruppen hadde jevnere skårer og skårene var generelt sett litt høyere enn den yngste aldersgruppen. Fordelingen for spørreskjemaet viste seg å være svært spiss, dette var grunnet at de fleste barna hadde skårer innenfor normalområdet noe som førte til en opphopning i den nedre del av skalaen.

Reliabilitetsanalysen viste at screeningtesten Språk 5-6 ikke har så god reliabilitet som en skulle ønske, i hvert fall om en ser på de enkelte deltestene. Summen av testen har reliabilitet som ligger over grensen, på .8, for hva som kan godtas. En grunn til den lave reliabiliteten kan være at det er et lite utvalg og at utvalget har for lite variasjon. Kanskje ville reliabilitetskoeffisienten blitt annerledes om en hadde hatt med barn med språkvansker i utvalget. Det at det er lite variasjon i utvalget gjorde at en del items ikke hadde varians og måtte utelukkes fra reliabilitetsanalysen. Dette reduserte antall items, noe som kan ha påvirket reliabiliteten.

Korrelasjonsanalysen viste en moderat korrelasjon mellom de to variablene ved bruk av Spearmans Rho. Denne var signifikant på 0.01 nivå som betyr at man med 99% sikkerhet kan stole på resultatet som et virkelig funn. Det ble også gjort en analyse av sammenhengen mellom variablene i krysstabellen. Her var det også en signifikant korrelasjon, men den var svak. Ved å gruppere resultatene for å forenkle krysstabellen, kan noe informasjon ha gått tapt. Det kan være en årsak til at korrelasjonen virker sterkere ved bruk av Spearmans Rho.

5.1 Diskusjon av oppgavene i testen

I deltesten setningsminne klarte *alle* barna de tre første oppgavene. Det er de setningene som skiller Språk 5-6 fra Språk 6-16. På én side kan det være en indikasjon på at de nye setningene er for enkle. På en annen side så er testens ment for å finne barn som strever med språket. Barn med store vansker må også mestre i det minste én oppgave for at en skal vite at oppgaven er forstått. Hvis barnet ikke mestrer noen oppgaver, heller ikke forsøksoppgaven, så er det vanskelig å vite om barnet ikke mestrer oppgaven på grunn av en språklig svikt, eller om det har å gjøre med at han ikke forstod hva han skulle gjøre.

Ved å se på frekvensen for hvor mange som svarte riktig og galt på de forskjellige items i deltesten ordspenn, ser det ut til at rekkefølgen ikke går fra enklest til vanskeligst for barn mellom 5 og 6 år. Der er spesielt ett item som skiller seg ut. Item nr.11 (*Grønn-Kan-Kutt-Vind*) er det flere som har klart enn både den distinkte ordrekken som kommer før denne og den ordrekken som rimer som kommer før item nr. 11. Kanskje det kunne gjøres en omorganisering av items på bakgrunn av data fra barn på mellom 5 og 6 år.

Motsetninger har vist seg å være en vanskelig oppgave for barna. Det var få som klarte mer enn de tre første oppgavene, hvor oppgaven presenteres i en setning. Her er det også noe usynkront i forhold til hvilke oppgaver barna klarer i forhold til rekkefølgen av spørsmålene. Blant annet er det to stykker som har fått riktig på item nr.7 (*Hva er det motsatte av synke?*) mens ingen klarte item nr.6 (*hva er det motsatte av tam?*), det var også tre som klarte item nr.8 (*hva er det motsatte av krympe?*) og to stykker klarte item nr.10 (*hva er det motsatte av stramme?*). Dette tyder på at rekkefølgen på items i denne oppgaven kunne vært revidert med tanke på den nye aldersgruppen.

Det kan virke som at deltesten begreper var spesielt problematisk for barna i denne undersøkelsen. Det var få av barna som oppnådde høye skårer på oppgaven motsetninger som er en del av deltesten begreper. En god del av barna forstod ikke

helt hva oppgaven gikk ut på og begynte å rime i stedet. Denne oppgaven ble beholdt uforandret fra Språk 6-16. Det at barna på 5-6 år fant denne oppgaven spesielt vanskelig tyder på at en form for tilpasning burde vært gjort. En kunne lagt til noen enklere oppgaver, eller kanskje burde alle items vært presentert i en setning. Eventuelt kan det være at motsetninger er for vanskelig for barn i denne alderen.

Ordkunnskap var også en vanskelig oppgave for barna i dette prosjektet. På den første oppgaven (*Hva er en katt?*) var riktige og gale svar fordelt 50/50, til sammenligning var det bare 6 av barna som ikke oppnådde et riktig svar på den tredje oppgaven (*hva er en telefon?*). Den første oppgaven er kanskje for vanskelig som en startoppgave. For at svaret skal være riktig skal barnet kategorisere en katt som et dyr, noe som kan være vanskelig for barn i denne alderen. En skal helt ned til oppgave 9 før antall riktige svar er lavere enn antall gale svar. Før dette var det betydelig flere riktige svar enn gale. Her ser det ut til at rekkefølgen på oppgavene fungerer for barna i denne studien også, men den første oppgaven er for vanskelig.

Gjennom analysen av barnas resultater i kapittel 4.1.1 viste det seg at det var noen forskjeller mellom aldersgruppene i denne studien. Det kan skyldes at barna er i stor utvikling i denne alderen og at det kan gi utslag i store forskjeller i språklige ferdigheter. Det at det viste seg å være en del forskjeller mellom aldersgruppene gjør at det kanskje ville være en idé å benytte egne normer for de to aldersgruppene.

Fordelingen for resultatene av testen viste seg å være tilnærmet normalfordelt. Noe som er en god forutsetning for videre analyser. Derimot viste fordelingen for resultatene fra spørreskjemaet å være svært spiss. Det er på grunn av at mange av barna har skårer innenfor normalområdet, slik at det blir en opphopning i den nedre del av skalaen. Det kan være at hvis en hadde hatt et større utvalg av barn, og da kanskje også barn med språkvansker så ville fordelingen blitt mer symmetrisk.

5.2 Diskusjon av reliabilitet

Som vist i kapittel 2 om teoretisk og empirisk referanseramme, er det viktig for validiteten at kriteriet som testen skal måles opp mot er reliabelt. Det er også essensielt at testen i seg selv er reliabel. Det blir derfor foretatt en vurdering av testens og spørreskjemaets reliabilitet som et ledd i vurderingen av testens validitet.

I følge Gall, Gall & Borg (2007) skal reliabiliteten være .8 for at testen skal kunne brukes i forskningssammenheng. Det viser seg at reliabiliteten til testen ikke er god nok. Det er bare én deltest for én aldersgruppe som oppnår god nok reliabilitet. Sumskårene oppnår derimot god nok reliabilitet. Men hva kan være grunnen til at reliabiliteten til Språk 5-6 er så mye dårligere enn reliabiliteten til Språk 6-16? Svaret kan ligge i utvalget, at et større utvalg som også inkluderer barn med språkvansker ville gi et annet utslag. Også med tanke på varians i resultatene kunne det være en forklaring da en del items ble utelukket fra analysene for Språk 5-6 på grunn av manglende varians. Antall items spiller en rolle for reliabiliteten, da mange items fører til bedre reliabilitet (Crocker & Algina, 1986; Gall, et al., 2007).

Deltestene skal derimot ikke brukes hver for seg, det blir sett på sumskåren. Dermed er det den som er den viktigste å se på. Sumskårene for hele utvalget og de to undergruppene oppnår kravet om en reliabilitetskoeffisient høyere enn .8.

Reliabilitetskoeffisientene for deltesten begreper er en del lavere enn de andre. Det kan være at usikkerhet rundt skåringen av begrepstesten har påvirket reliabiliteten, da skåringen innebærer en viss skjønnsmessig vurdering av den enkelte testleder. Da testlederne ikke hadde noen fastsatt skåringsmal for de nye oppgavene i testen kan det forekomme ulikheter i skåringen.

For å se om reliabiliteten kan godtas kan en sammenligne reliabiliteten til Språk 5-6 med reliabiliteten til andre lignende tester. For eksempel har CELF – preschool også relativt lav reliabilitet for enkelte deltester. For barn mellom 5,0 og 6,0 år har CELF – Preschool reliabilitetskoeffisienter ned mot 0.30 for noen deltester (Wiig, Secord, & Semel, 1992). Samlet for alle aldersgrupper har noen deltester

reliabilitetskoeffisienter ned mot 0.50. CELF- Preeschool har , i likhet med Språk 5-6, tilfredsstillende reliabilitet for sumskåren (Wiig, et al., 1992). En ser at også andre tester for førskolebarn har lignende eller lavere reliabilitet, slik at reliabiliteten til Språk 5-6 kan aksepteres.

Det finnes også tall på reliabilitet fra tidligere på disse instrumentene, som ble presentert i kapittel 2.8.1. Testen er ikke helt den samme, da den nye testen Språk 5-6 har noen enklere oppgaver i begynnelsen av noen av deltestene.

Reliabilitetskoeffisientene er høyere for Språk 6-16 enn de er for Språk 5-6. Dette kan ha noe med størrelsen, og variansen i utvalget som ligger til grunn for reliabilitetsanalysene å gjøre og aldersgruppen som er med i det enkelte utvalget. Den største forskjellen ligger i deltesten begreper. Her har Språk 5-6 en reliabilitetskoeffisient på .642 mens Språk 6-16 har en reliabilitetskoeffisient på .90. Denne forskjellen er ganske stor, og kan kanskje forklares av at vokabularet er i stor utvikling for barna mellom 5 og 6 år. Dermed er det kanskje større variasjon i deres vokabular, mens de barna som er eldre enn 6 år kanskje har et mer konsistent vokabular. Innen et barn har blitt 6 år har det lært i overkant av 14000 ord, hvis man tar med i beregningen at vokabularet ikke vokser alvorlig mye før 18 mnd vil det si at barnet lærer ca 9 ord om dagen i gjennomsnitt (Carey, 1978). Kanskje kan forskjellen mellom barna opptil 6 år og de som er eldre enn 6 år delvis forklares ved at vokabularet til barna under 6 år er i en enorm vekst noe som også fordrer store forskjeller, og dermed store ulikheter i skårer på oppgaver som tester vokabular. Kanskje er de items som er lagt til på begynnelsen av deltestene for enkle. Kanskje er det ikke så stor forskjell mellom barna på opp til 6 år og de som er fylt 6 år at det er behov for så enkle items i begynnelsen av hver deltest. Dette kan støttes av det faktum at i flere av deltestene hadde alle barna fått riktig på de første itemsene slik at disse måtte utelukkes fra reliabilitetsanalysene. Kanskje har dette spilt inn på reliabiliteten til de enkelte deltestene og testen som helhet. En annen mulighet en må vurdere er at det utvalget som ble testet i dette prosjektet bestod av barn som man antok ikke hadde vansker. Det kan være at barn som faktisk har vansker trenger disse

enkler oppgavene. Selv om barn med vansker også skulle få alt riktig på de enkleste items, så er det viktig for dem å oppleve mestring. Som nevnt tidligere skal en screeningtest ikke gjennomføres hvis det vil være til skade for barnet. Det å oppleve at en ikke mestrer en eneste oppgave på en test kan være en påkjenning som kan påvirke barnets selvfølelse (Skaalvik & Skaalvik, 1996). Da er det testleders ansvar å sette seg inn i testen og bli kjent med barnet og ta en avgjørelse på forhånd om det å ta denne testen vil være til skade eller til gagn for barnet.

De andre reliabilitetskoeffisientene er også en del høyere for Språk 6-16 enn de er for Språk 5-6. Det kan kanskje tyde på at testen Språk 5-6 egner seg dårligere for bruk på disse yngre barna enn Språk 6-16 gjør for barn mellom 6 og 16 år. Om en ser på forskjellen mellom barn på 5-5,5 år og de på 5,5 – 6 år så ser man også en forskjell i den retning at de eldre barna har mer konsistente svar og dermed høyere reliabilitetskoeffisient.

En ser samme trend med tall fra spørreskjemaet som en gjør med screeningtesten. Ottem presenterte sine tall fra valideringen av skjemaet i Skolepsykologi nr.1 2009 (Ottem, 2009). Reliabilitetskoeffisientene fra den studien ligger litt høyere enn de som har kommet frem gjennom analyser av materialet i dette prosjektet, men det er ikke så mye. Det er noen forskjeller mellom Ottens utvalg og utvalget av barn i dette prosjektet. Ottem hadde med både normalfungerende barn og barn med språkvansker. Han hadde også et større antall barn og disse barna var eldre enn de som er med i dette studiet. Dette kan være med på å forklare forskjellene. De som svarte på spørreskjemaene i Ottens studie var lærere. Lærere og førskolelærere har ulik utdanning og omgangsmåte med barna slik at de stiller med noe ulike forutsetninger for å svare på skjemaet. Lærernes hverdag og omgang med barna er preget av at de har et pensum som skal læres og forskning vider at mye av lærernes hverdag går med til blant annet å få ro og orden i klasserommet (Strøm, Borge, & Haugbakken, 2009). Barnehagen er mer preget av lek og moro enn skolen er og dermed får førskolelærerne en annen type omgang med barna enn lærerne får med elevene i skolen. En lærer har derimot flere holdepunkter for hva barna skal kunne,

og kan dermed lettere legge merke til de som ikke mestrer det som er forventet at de skal kunne. Disse faktorene kan være bidragsytende til mer eller mindre konsistente og dermed reliable svar fra lærere og førskolelærere. Et viktig poeng er at både med lærere og med førskolelærere som respondenter på skjemaet, så oppnår ”20 spørsmål om språkferdigheter” svært reliable svar som gir godt grunnlag for valide slutninger.

5.3 Diskusjon om testens validitet

Når en har fått etablert at måleinstrumentene er reliable nok til å brukes i forskning kan man gå videre til å vurdere validiteten til slutningene som dras fra resultatene. Oppgavens problemstilling søker å si noe om sammenhengen mellom de to instrumentene som ble benyttet i dette prosjektet. Denne sammenhengen undersøkes for så å kunne si noe om den nye testens gyldighet som instrument for å måle barns virkelige språklige ferdigheter. Det kan argumenteres for at spørreskjemaet er et mer økologisk valid instrument i utgangspunktet i og med at det er en hverdagsvurdering fra førskolelæreren som ligger til grunn. På den måten blir ikke barnet bevisst på at det blir vurdert, noe som igjen ikke vil påvirke barnet. I en testsituasjon kan barnet bli påvirket av oppgavene og situasjonen på en måte som gjør at han ikke klarer å vise sine virkelige språklige ferdigheter. Derfor var det interessant å se på om det fantes noen sammenheng mellom et økologisk valid instrument og et som kunne være påvirket av situasjonen.

Gjennomgangen av utviklingen i synet på validitet i kapittel 2.4 munnet ut i at det nåværende synet inneholder fem vurderinger av validitet som baserer seg på innholdet i testen, responsprosess, indre struktur, relasjoner til andre variabler og konsekvenser av testing. Disse vurderingene vil nå bli gjort av screeningtesten Språk 5-6.

5.1.1 Vurdering av testens validitet basert på innholdet i testen.

Som nevnt i kapittel 2.4 omhandler dette om oppgavene i testen representerer det domenet til fenomenet som skal studeres. Når en test skal konstrueres tas en del valg. En velger hvilken type oppgaver en skal bruke, i dette tilfellet minneoppgaver og

oppgaver som undersøker barnets begreper. Deretter er det valgt om det skal være tall, ord, nonord eller bilder som skal huskes. I denne testen er det valgt setninger og ordrekker. Oppgaven setningsminne er ikke et rent mål på fonologisk minne da en kan dra nytte av langtidsmminnet for å huske sammenheng i setningen (Ottem & Frost, 2005). Minne for tall og nonord er mye brukt i oppgaver som søker å undersøke det fonologiske minnet (Alan Baddeley, et al., 1998; Gathercole & Baddeley, 1990). Tall kunne vært problematisk å benytte da barna som testes med Språk 5-6 er relativt unge og har begrenset kjennskap til tall, det er også vanskelig å vite hvem som eventuelt ville få matematikkvansker senere. Forskning har vist at matematikkvansker kan spille inn på evnen til å huske tall (Hulme & Roodenrys, 1995). En forklaring på dette kan være at tilgangen til de fonologiske representasjonene av tallord i langtidsmminnet er tregere enn den for andre ord hos barn med matematikkvansker (Hulme & Roodenrys, 1995). Dermed kan barna ha uoppdagede matematikkvansker som spiller inn på deres evne til å huske tall, som vil ha følger for deres skåre på testen og som kan regnes som systematiske målefeil. Systematiske målefeil er en trussel mot validiteten (Kleven, 2002b). Basert på forskning om det fonologiske minnet og hva det har å si for begrepslæringen er det inkludert en todelt oppgave om begreper i testen (Alan Baddeley, et al., 1998).

5.1.2 Vurdering av testens validitet basert på responsprosess

Ved denne typen vurdering må en, som tidligere nevnt, undersøke om den typen responser som kreves av barnet passer til det begrepet som skal undersøkes. Oppgavene i screeningtesten Språk 5-6 krever responser i form av tale. Et eventuelt problem med dette kan være hvis det at barnet har lavt taletempo påvirker prestasjonen på minneoppgavene. Hulme og Tordoff (1989) foreslo ut fra sine resultater at kapasiteten i det fonologiske lageret vokste med alderen på grunn av økt taletempo. Barn i 5-6 årsalderen er i stor utvikling, noe som fører til store ulikheter, også i taletempo. Erfaringsmessig var det ikke noen som hadde svært lavt taletempo i forhold til noen andre, men noen barn brukte svært lang tid på å tenke før de gjenga ordrekkene til testleder, noe som kanskje kan ha påvirket deres prestasjoner. En

alternativ metode kunne være å bruke bilder i denne oppgaven, slik at barnet skulle respondere ved å peke på bildene i riktig rekkefølge. Bruk av bilder ville eliminere et eventuelt problem med barnets taletempo. Det skulle være mulig å bruke bilder i både Setningsminne og Ordspenn. Ved bruk av bilder ville en også inkludere den visuospatiale skisseblokken, der ville bildene kunne bli ”oversatt” til tale og slik ført inn i den fonologiske løkken.

Screeningtesten Språk 6-16 har blitt nøye utprøvd og oppnådd relativt god validitet. Responsprosessen i denne er lik den i Språk 5-6, den eneste forskjellen er aldersgruppen den skal brukes på og at det har blitt lagt til noen enklere items i begynnelsen av noen av deltestene. Språk 5-6 er førskoleutgaven av Språk 6-16, de to testene har ulikt aldersspenn de skal favne om. Selv om responsprosessen har fungert for aldersgruppen 6-16 år så er det ikke sikkert at den fungerer like godt for aldersgruppen 5-6 år. Den største forskjellen mellom disse gruppene er nok at barn fra 6-16 år går på skolen og er vant til skolens arbeidsmåte, mens barn mellom 5-6 år fremdeles går i barnehagen hvor hverdagen er mer preget av lek. Barnehagebarn er kanskje ikke så vant til å sitte stille å svare på spørsmål, slik som i en testsituasjon. Det at situasjonen er fremmed for dem kan påvirke deres prestasjon. Dette er eksempel på når en økologisk validitetstekning kan være nyttig. I et økologisk validitetsperspektiv ville en forsøke å gjøre testsituasjonen lik en naturlig situasjon hvor barnet bruker språket sitt. Hva som er en naturlig situasjon er forskjellig fra person til person, noen er utadvent og snakker med hvem som helst, mens andre er mer reserverte. Derfor kan det å bruke spørreskjemaet ”20 spørsmål om språkferdigheter” være et godt supplement til tester. Det vil gi et annet perspektiv på barnets språkferdigheter, eventuelle uenigheter mellom spørreskjemaet og testen kan diskuteres mellom testleder og førskolelærer.

5.1.3 Vurdering av testens validitet basert på testens indre struktur

Som nevnt i kapittel 2.4 handler dette om å se hvorvidt de underliggende komponentene som testen undersøker, fonologisk minne og begrepsmessig utvikling, er viktige faktorer med tanke på språkvansker. Språkvansker er et stort begrep, for å

begrense det så tenkes det her på spesifikke språkvansker. Språket er et komplekst system og grunnen til at enkelte har vansker med det er ikke mindre kompleks. Det er vanskelig å svare på om ”løsningen” med at en svikt i det fonologiske minnet kan gjøre den begrepsmessige utviklingen vanskelig og at det spiller en viktig rolle i spesifikke språkvansker er den riktige løsningen. Det finnes mange forskjellige synspunkter på hva som ”forårsaker” språkvansker. Det er vanskelig å lage en test som tar hensyn til alle de forskjellige synspunktene. Noen forskere har vist at det er en sammenheng mellom en svikt i det fonologiske minnet og språkvansker (Gathercole & Baddeley, 1990). Denne svikten skal ligge til grunn for at barn med språkvansker skal ha vansker med oppgaver som ordspenn. Setningsminne er som tidligere nevnt en oppgave som ikke bare benytter det fonologiske minnet, men som sannsynligvis også tar i bruk noe av langtidsmminnet. Deltesten begreper måler barnets begrepsmessige utvikling. Ut fra den antakelsen som ligger til grunn for testen, burde en svikt i det fonologiske minnet føre til at barnet har vanskelig med å lære nye ord som igjen vil føre til en svikt i den begrepsmessige utviklingen. På bakgrunn av det grunnlaget som testen er konstruert på så vurderes testens indre struktur å være tilfredsstillende.

5.1.4 Vurdering av testens validitet basert på dens relasjoner til andre variabler

Da dette er en korrelasjonsstudie er dette den vurderingen av testens validitet som vil bli lagt mest vekt på. I hovedprosjektet har de data fra flere tester som den nye screeningtesten vil bli målt opp mot, men i dette delprosjektet er det bare testens forhold til spørreskjemaet ”20 spørsmål om språkferdigheter” som vil bli undersøkt. Forholdet mellom spørreskjemaet og testen er et uttrykk for konvergent og samtidig validitet, da de begge undersøker former for språklige ferdigheter og data er samlet inn på omtrent samme tid.

Her kan det settes opp en nullhypotese som lyder som følger: Det er ingen sammenheng mellom screeningtesten Språk 5-6 og spørreskjemaet ”20 spørsmål om språkferdigheter”. Målet er da å forkaste denne hypotesen da det er ønskelig med en

sammenheng for å kunne si at testen måler barnets virkelige språkferdigheter, slik de fremstår i en naturlig situasjon.

I kapittel 4.2.2 ble det presentert at det fantes en moderat negativ korrelasjon mellom de to instrumentene. Korrelasjonen var signifikant på 0,01 nivå. Det er naturlig at korrelasjonen er negativ da en *lav* skåre på spørreskjemaet er et tegn på gode språklige ferdigheter, mens en *høy* skåre på testen tyder på gode språklige ferdigheter. Signifikansnivået forteller at det er $< 1\%$ sjanse for at sammenhengen er en ren tilfeldighet. Det vil si at det er $< 1\%$ sjanse for at det har blitt begått en type 1 feil (feilaktig avvisning av nullhypotesen) (Connolly, 2007; Tabachnick & Fidell, 2001). Det er relativt trygt å forkaste nullhypotesen og anta at det virkelig finnes en sammenheng mellom de to instrumentene.

Gjennom krysstabellen i kapittel 4.2.3 kom det frem at $p=.002$ for krysstabellen. p er også her lav nok til at vi kan forkaste nullhypotesen og en kan anta at det er en reel sammenheng mellom de to måleinstrumentene. Det er 0,2 % sjanse for at det er begått en type 1 feil og at nullhypotesen faktisk er sann. Denne muligheten er svært liten og det er derfor rimelig å feste lit til de resultatene man har funnet i denne undersøkelsen. Styrken på sammenhengen mellom instrumentene en fant gjennom bruk av krysstabellen kan karakteriseres som svak. Den er svakere enn den en fant gjennom korrelasjonsanalysen, når en ikke grupperte skårene. Det kan skyldes at en mister noe informasjon når svarene blir gruppert, og at Cramers V og Spearmans Rho ikke er like mål. Når en ser på fordelingene i krysstabellen så ser man at det virker som at det er relativt stor enighet om de forskjellige barna mellom spørreskjemaet og testen. Det er få store avvik.

Ottem skal lage skåringskriteriene for den nye testen og da vurdere hvilken cutoffskåre som er hensiktsmessig for Språk 5-6, det skal gjøres på bakgrunn av et større tallmateriale enn det har vært tilgang til i tilknytning til denne oppgaven. I dette prosjektet har det derfor blitt benyttet den cutoffskåren som er satt for Språk 6-16 (Ottem & Frost, 2005) og som Ottem (2009) brukte for spørreskjemaet da han validerte det. Men det vi ser nå er at både testen og spørreskjemaet finner noen i

utvalget vårt som har vansker, men det dreier seg ikke alltid om de samme. Som vist i kapittel 4.2.3 så er det ikke alle klassifiseringene av barna som de to instrumentene er enige om. Ottem og Frost (2005) karakteriserer barn som skårer $-1,5$ standardavvik eller dårligere på screeningtesten Språk 6-16 som å være i en risikogruppe for språkvansker. Det er den samme inndelingen som er brukt på materialet fra denne undersøkelsen. Med denne cutoffskåren er det 8 barn som faller i risikogruppen i følge testen. Med utgangspunkt i eventuell diagnostisering av spesifikke språkvansker så kan cutoffskåren senkes noe. Forskere er uenige om hvor man bør sette grensen for når en vanske skal kunne sies å være en spesifikk språkvanske. Det ligger mange aspekter til grunn for å gi en slik diagnose, men en av dem er at barnet skal skåre betydelig lavere enn gjennomsnittet på standardiserte språktester (Bishop, 1997; Leonard, 1998). I følge Leonard (1998) så er kravet at barnet skal skåre $-1,25$ lavere enn gjennomsnittet. Hvis en setter cutoffskåren på $-1,25$ SD under mean, finner testen 15 barn som er i risikosonen for språkvansker. Det tilsvarer 12,7% av utvalget i dette prosjektet. Det er et svært høyt tall i forhold til at Leonard (1998) anslår forekomsten av spesifikke språkvansker til å være 7 %. Det er lite trolig at andelen barn i utvalget som har språkvansker er så høyt. Det tyder på at et lavere cutoffpunkt enn $-1,25$ SD er passende for denne testen.

Med utgangspunkt i en cutoffskåre på 31, som Ottem satte for "20 spørsmål om språkerfardigheter" (Ottem, 2009), "finner" spørreskjemaet 16 barn som har vansker med språket i større eller mindre grad. Av disse er det bare 3 barn som blir satt i den høyeste vanske kategorien av førskolelærerne, de får en skåre på 51 eller høyere. Ut fra testresultatene er det 15 barn som har vansker med språket, med et cutoffpunkt på $-1,25$ SD, av de 15 er det 8 som blir satt i den gruppen med mest vansker. Testen har en grovere skala for vansker enn spørreskjemaet. Den har etter Ottem og Frosts (2005) gradering bare én vanskegruppe ($\leftarrow -1,5$), men i denne oppgaven har det blitt utvidet til to grupper ($\leftarrow -1,5$ og $-1,5$ til -1) for å få med cutoffskåren for diagnostisering av spesifikke språkvansker. Testen har da tre grupper i normalområdet, mens spørreskjemaet har to grupper for normal fungering og tre for vansker. Testen har da funnet 15% (18 stk) som strever med språket, og

spørreskjemaet har funnet 13,5% (16 stk) som strever med språket. Andelen barn som disse instrumentene finner at strever med språket er relativt høy. Som nevnt opererer Leonard (1998) med at ca 7 % av alle barn har spesifikke språkvansker. I lys av det kan det virke lite trolig at så mange av barna i utvalget har vansker med språket. De høye tallene kan skyldes en unøyaktighet i instrumentene. Kanskje det er vanskelig for førskolelærerne å vurdere hva som er normalt for barn i denne aldersgruppen å mestre. Kanskje bruker førskolelærerne de flinkeste barna som mal og vurderer alle de andre barna opp mot dem. Det kan bli en urettferdig vurdering da de flinkeste barna kanskje er eksepsjonelt flinke, og målt opp mot disse vil selv normale barn komme ut med vansker. En mulighet for å motvirke dette ville være om spørreskjemaet hadde hatt en skala som også gav mulighet for å merke de som var eksepsjonelt flinke. Nå starter skalaen på ”ingen vansker”, men det er store forskjeller også innad i denne gruppen.

Selv om spørreskjemaet bare ”finner” ett barn mer enn testen, betyr ikke det at de to instrumentene er enige om hvem som har språkvansker. Der er noen barn som instrumentene absolutt ikke enes om. En forklaring på hvorfor de ikke er enige, kan kanskje finnes i skalaen ”relasjoner til andre” i spørreskjemaet.

Enkelte case sett i sammenheng med ”relasjoner til andre”

Skalaen ”relasjoner til andre” finnes i spørreskjemaet og kommer i tillegg til de 20 spørsmålene om språkferdigheter. Skalaen er lagt til for å kunne inkludere atferdsaspektet i en utredning av barna. Vansker med atferd kan både være en isolert vanske og en vanske som er en følge av språkvansker (Botting & Conti-Ramsden, 2000). I dette avsnittet vil noen barn som instrumentene ikke er enige om bli presentert og en forklaring vil søkes i skalaen ”relasjoner til andre”.

Ett eksempel er et barn som skårer 11 poeng på Språk 5-6, får 20 poeng på spørreskjemaet av førskolelæreren, og får 10 poeng på skalaen ”relasjoner til andre”. 11 poeng på Språk 5-6 er den laveste skåren i utvalget og tilsvarer en skåre som ligger 2,37 SD under gjennomsnittet. På spørreskjemaet ”20 spørsmål om språkferdigheter” tilsvarer en skåre på 31, som er cutoffskåren, 31 % av maksimal

skåre. 10 poeng på ”relasjoner til andre” tilsvarer 28,6 % av maksimal skåre som ligger nært opp til den kritiske grensen for når en vil si at barnet har vansker, men befinner seg fremdeles i normalområdet. Her er det et misforhold mellom hva resultatet på testen sier og hva skåren på spørreskjemaet sier, ved at den som har fått den dårligste skåren på Språk 5-6 har fått en skåre som tilsier normal språklig fungering på spørreskjemaet. I tillegg kan en si at barnet har fått en skåre på skalaen ”relasjoner til andre” som indikerer at han har fått utslag på noen utsagn, men poengsummen er fremdeles innenfor normalområdet. Det kan være at testen har fanget opp noen vansker som ikke er lett å få øye på for førskolelæreren.

Førskolelæreren har registrert *noe* i barnets relasjoner til andre, da barnet er bare 1 poeng fra å havne over cutoffskåre for atferdsvansker. Det kan være at de vanskene med språket som barnet har ikke er så synlig for førskolelæreren, men de kan gjøre det vanskelig for barnet i relasjoner til andre barn. Det kan også være at de vanskene som testen finner ikke er reelle, selv om dette virker lite trolig da barnet oppnår en så lav skåre. En annen mulighet er at barnet har blitt påvirket av testleder og testsituasjonen og ikke klarer å vise sine sanne ferdigheter. Barnet kan ha misforstått oppgavene og på grunn av det fått en dårlig skåre, eller han kan ha hatt en dårlig dag.

Et annet eksempel er to barn som er plassert i øvre normalområdet av testen (0 til +1 SD) blir satt i (31-40 %) og (41-50 %) av førskolelærerne. Det er vanskelig å vite hvorfor instrumentene ikke er samstemte i kategoriseringen av disse to barna men en mulig forklaring kan kanskje finnes om man ser på hvilke skårer disse barna fikk på ”relasjoner til andre”. Disse skårer 10 og 11 på ”relasjoner til andre”. Det er 28,6 % og 31,4 % av maksimal skåre, de ville blitt karakterisert som å ligge over, eller nær grensen for atferdsvansker. Det kan være at de språkvanskene som førskolelæreren registrerer hos disse barna egentlig er atferdsvansker, eller det kan være vansker med språket som faller utenfor det som testen undersøker.

Oppsummering av diskusjon vurdering av testens validitet basert på dens relasjon til en kriterievariabel.

Det er en signifikant sammenheng mellom testen og spørreskjemaet, korrelasjonsanalysen viser en moderat sterk sammenheng. I krysstabellen er

resultatene gruppert, sammenhengen mellom variablene er da svak. Det kan komme av at grupperingen av resultatene gjør at man mister informasjon som gjør sammenhengen svakere. Det at sammenhengene ikke er sterke, men signifikante, tyder på at utredning av språkferdigheter hos barn er et område som krever ulike metodiske tilnærminger. Screeningtesten Språk 5-6 tar for seg en bestemt del av språklige ferdigheter, men det finnes også mange andre. De oppgavene som er valgt ut i testen er bare et utvalg av hele domenet, og kan dermed ikke sies å være uttømmende. Spørsmålene i listen ”20 spørsmål om språkferdigheter” er også valgt ut fra et univers av mulige spørsmål. Det vil si at det finnes en enorm mengde av alternative spørsmål som kunne stilles. En vet også at et spørsmål kan oppfattes forskjellig, dermed kan det svaret en får være svar på ulike spørsmål avhengig av hvordan spørsmålet ble oppfattet.

Med dette i tankene vil kanskje det å kombinere de to instrumentene være en reell mulighet. De vil kunne gi ulike typer opplysninger. Testen ville gi resultater i form av tall som lett kan sammenlignes med normene, mens spørreskjemaet kan gi informasjon utover det man får fra testen og fra et annet perspektiv.

Bishop (1997) etterspurte utredning av spesifikke språkvansker som omfattet både en testbasert diagnose og en nedtegning av hvordan barnets vansker påvirket barnet i hverdagen. Slik ønsket hun å rette fokus mot hvilke konsekvenser kommunikasjonsvanskene, som følge av spesifikke språkvansker, kan få for barna i deres hverdagsliv. Bishop (1997) skriver videre at hun har inntrykk av at foreldre og lærere har lett for å oppdage kommunikasjonsvansker som fører til at barna blir vanskelig å forstå eller har umoden setningsstruktur. Slike kommunikasjonsvansker er lettere å oppdage ved at de er hørbare. Derimot skriver Bishop (1997) at hun tror at foreldre og lærere har vanskeligere for å oppdage språkproblemer som påvirker forståelse, vokabularstørrelse og verbalt minne. Forskningen til Ottem (2009) motviser på et vis dette ved at de viser at lærerne i den undersøkelsen var svært enige i de karakteristikken som testen fant og testen var en test som i likhet med Språk 5-6 tester underliggende funksjoner som fonologisk minne og begrepsmessig utvikling.

Det er en signifikant sammenheng mellom spørreskjemaet og testen, denne sammenhengen er moderat. Om en hadde inkludert barn med påviste språkvansker i utvalget ville en kanskje fått sterkere sammenhenger mellom skjemaet og testen. Før en kan trekke noen sikrere slutninger om hvorvidt testen kan sies å ha tilfredsstillende validitet, ville det være nyttig å utvide utvalget til å omfatte barn med påviste språkvansker. Det ville blant annet kunne gi svar på om de enkleste oppgavene i testen slik den er nå er enkle nok for barna som har store språkvansker. Hvis et barn ikke mestrer noen oppgaver og heller ikke instruksjonsoppgaven, vil det være et problem for reliabiliteten. En ville da ikke vite om barnet ikke klarte oppgavene fordi han har store vansker på det området som oppgaven undersøker eller om han bare ikke forstod oppgaven og hva som forventes av ham for å løse oppgaven på en tilfredsstillende måte.

En av grunnene til at man er interessert i å vite noe om testens validitet er for å kunne si noe om hvor gyldige de slutningene som dras fra testens resultater er. For at en slutning skal være gyldig må den bygge på nøyaktige målinger (Murphy & Davidshofer, 2001).

5.1.5 Vurdering av testens validitet basert på konsekvenser ved testing.

For å gjøre en slik vurdering av testens validitet må en se på forutsette og uforutsette, positive og negative konsekvenser ved testing. Det å gjennomføre en utredning bygger på en tanke om at dersom barnet har vansker kan en få hjulpet barnet på best mulig måte. For å vite hvordan en skal hjelpe barnet, må en vite hva slags vansker det dreier seg om og gjerne hvilke sterke sider barnet har, slik at man kan bygge på dem. At barnet får hjelp for sine vansker er en positiv konsekvens av testing. En negativ konsekvens kan være dersom testens beskaffenhet gjør at barnet ikke klarer å prestere sitt beste og kommer ut med et dårligere resultat, som ikke gjenspeiler barnet virkelige ferdigheter. De som konstruerer en test har ansvar for å validere den og lage gode og forståelige retningslinjer for bruk og skåring av testen. Testkonstruktører bør også angi hvilken aldergruppe testen er egnet for. En som skal ta i bruk en test bør

være i stand til å vurdere om testen egner seg for å undersøke det aktuelle barnets ferdigheter. Det innebærer å være kjent med testens oppbygning og oppgaver, samt å vite hva testen undersøker og hvilken teoretisk referanseramme som ligger bak. Ved en slik vurdering sitter testleder med ansvaret for at testen blir brukt på korrekt måte og til et formål som den er gyldig for.

Konsekvenser av testingen kan også innbefatte tolkningene av skårene. En slutning fra testskårer som ikke er valide kan få negative konsekvenser. En slik ”ikke valid” tolkning kan oppstå dersom testleder ikke er kjent med tolkningskriteriene for testen, eller om testens skåringsmal eller tolkningskriterier ikke er tydelige nok.

Testresultater kan brukes og tolkes på flere forskjellige måter. Det må gjennomføres en validering for hver intendert bruk av testresultatene (AERA, et al., 1999). Om testresultater brukes på en annen måte enn det de er validert for så vil de ikke være gyldige, det kan få konsekvenser for barnet som har blitt testet. Disse konsekvensene kan være i form av feiltolkninger av resultater, noe som kan få negative følger for barnet.

Det er tilsiktet at testing skal føre til positive konsekvenser for barnet det gjelder. Utilsiktede positive konsekvenser er også velkomne. Man må forsøke å minske risiko for at testing skal få negative konsekvenser. En kan da forsøke å bruke tester med tilstrekkelig validitet, sette seg godt inn i manualen til testen, utføre og skåre den slik det er intendert fra testkonstruktøren. Deretter vil det å sørge for at slutningene en tar fra testresultatene er godt begrunnet være en sikkerhet. For at en skal kunne ta så valide slutninger som mulig og dermed minske risikoen for negative konsekvenser av testing, er det viktig at målingene testen gjør er nøyaktige.

5.4 Diskusjon av testens nøyaktighet.

Tidligere har det blitt nevnt at konsekvenser av testing i de siste årene har blitt tatt med i validitetsdebatten. Der spiller testens nøyaktighet en viktig rolle. Er testen unøyaktig og det derfor dras mange feilaktige slutninger vil en ende opp med over- eller underrapportering av vansker. Disse kan justeres ved å justere den kritiske

grensen, en må da tenke over hva som vil gjøre mest skade: Falsk-positiv eller falsk-negativ. I en metastudie som omhandlet overrapportering av vansker og hvorvidt dette var et problem, fant Glascoe (2001) at de barna som ble falsk- positive skårer vesentlig lavere på instrumenter som måler intelligens, språklige – og akademiske ferdigheter i motsetning til de som ble sann-negative. Dette kan tyde på at de barna som blir overrapportert for vansker kanskje vil ha nytte av den ekstra oppfølgingen eller uttreningen som kan være en følge av å bli identifisert som positiv for vansker. Glascoe (2001) fant også at overrapporterte barn hadde flere psykososiale risikofaktorer enn de barna som ble sann-negative, blant annet hadde foreldrene lavt utdanningsnivå og mange var etniske minoriteter. Med dette som argument kan en eventuelt anbefale å nedjustere cutoffskåren slik at flere blir identifisert som positive for vansker. I analysen så man at en cutoffskåre på -1.25, som er en del av kriteriet for å diagnostisere SSV, ikke passet til denne testen. Den cutoffskåren førte til at relativt mange barn, i et utvalgt hvor man antok at de som var med ikke hadde språkvansker, ble kategorisert som positiv for vansker. På spørreskjemaet var det flere barn som ble kategorisert som å ha vansker enn på testen, men vanskeskalaen for spørreskjemaet er mer findelt og en kan da diskutere om cutoffskåre for barn på 5-6 år bør være høyere enn for eldre barn. Det man risikerer med underrapportering av vansker er at hvis et barn først blir screenet med en test og får et resultat som viser at barnet er negativt for vansker vil terskelen for å gjenoppta en eventuell uttrening av barnet, for de samme eller lignende vansker, ligge høyere. Det kan føre til at barn blir oversett og ikke får den hjelpen de trenger, eller at det tar lang tid før vanskene blir oppdaget.

Testens nøyaktighet måles ved å se på testens evne til å skille mellom de som har språkvansker og de som ikke har det. Forskere mener at sensitiviteten til testen bør være over 70 % og spesifisiteten mellom 70 og 80 % (Glascoe, 1997; Rydz, et al., 2005). Ut fra de analysene som er gjort på det datamaterialet som var tilgjengelig i dette prosjektet fant man at Språk 5-6 har en sensitivitet på ca 43,8 %. Testen er dermed altfor lite nøyaktig i å finne frem til hvilke barn som har språkvansker. Analysene viser at testen har en sensitivitet på 89,2 %. Testen har da en mer enn god

nok evne til å sile ut de barna som ikke har vansker med språket. Det at sensitiviteten er så lav kan ha noe å gjøre med at utvalget er relativt lite for denne typen analyse. De fleste barna, 77,1 %, havnet i sann negativ gruppen og dermed vil de andre gruppene inneholde få barn. Når en har et lite tallmateriale vil enhver feil ha store konsekvenser. Utvalgskriteriene fordret et utvalg av barn uten antatte språkvansker, dette kan ha gjort sitt til at utvalget bestod av en relativt homogen gruppe. Kanskje ville testen vise seg å være mer nøyaktig om en hadde hatt en gruppe barn med påviste språkvansker, eller hvis en gjorde en prediktiv studie.

Gjennom utprøvingen av Språk 6-16 fant Ottem og Frost (2005) at 89 % av barna med påviste språkvansker ble riktig klassifisert og 84,7 % av de med normal språkutvikling ble riktig klassifisert av testen. Til motsetning ble bare 6 % av de med språkvansker riktig klassifisert av Språk 5-6 (sann positiv) og 77 % av de med normal språkutvikling ble klassifisert riktig av Språk 5-6 (sann negativ). Kontrastene her er store, spesielt for de barna med språkvansker. Resultatene for Språk 5-6 kan muligens forklares av at det ikke bevisst er inkludert barn med språkvansker i utvalget. Et lite tallmateriale kan føre til at små feil får store utfall.

Ut fra den analysen av testens nøyaktighet som er gjennomført her finner en ikke at testen er nøyaktig nok til å finne de med vansker, men at den er god nok til å ekskludere de uten vansker fra en vanskegruppe. Dette kan nok ha noe med å gjøre at det utvalget det har vært tilgang til i denne undersøkelsen i hovedsak består av barn som antas å ikke ha noen form for vansker. Analysen blir dermed ikke gjort på et ideelt grunnlag. På bakgrunn av de data som foreligger, og med tanke på at utvalget kanskje er noe homogent, så kan man si at testen har god nok spesifisitet, sensitiviteten er derimot langt fra god nok. Denne analysen må dog tolkes med forsiktighet da en helst skulle sett at den ble gjennomført med en gruppe av barn med påviste språkvansker, og som en del av en longitudinell studie.

5.5 Oppsummerende vurdering av testens validitet

Sammenhengen mellom spørreskjemaet og testen er signifikant. Korrelasjonen mellom de to variablene er moderat. Dette kan kanskje skyldes utvalget. En gruppe med barn som har påviste språkvansker skulle vært med i studien. Da ville en kunne få undersøkt hvordan testen kategoriserte disse som man visste hadde vansker. Disse resultatene taler for en flermetodisk tilnærming til utrening av språkvansker.

Testens nøyaktighet er ikke helt slik man skulle ønske. Den identifiserer godt nok de uten vansker, men finner ikke godt nok de som har vansker. Dette skyldes nok utvalget. Det burde være to grupper, en med og en uten vansker og studiet burde være longitudinelt. Tolkningene fra denne analysen må tas med forsiktighet da datamaterialet ikke er ideelt for å undersøke testens nøyaktighet.

6. Avsluttende kommentarer

Intensjonen med dette delprosjektet var å bidra til valideringen av den nye screeningtesten av barns språk, Språk 5-6. To variabler ble benyttet, den nye screeningtesten og spørreskjemaet ”20 spørsmål om språkferdigheter”.

Problemstillingen for oppgaven var: *Hvilket forhold er det mellom førskolelæreres vurdering av barns språklige ferdigheter og disse barnas resultater på Språk 5-6?*

For å svare på denne problemstillingen ble to forskningsspørsmål formulert:

3. Måler testen det den skal måle?

Dette spørsmålet handlet om kriterievaliditet. Hvor det ble antatt at testen søkte å si noe om barns språklige ferdigheter. Spørreskjemaet ”20 spørsmål om språkferdigheter” ble brukt som kriterium på dette. Spørreskjemaet var allerede validert av Ottem (2009) og kan betraktes som en mer økologisk valid måte å måle språklige ferdigheter på, da informasjonen som innhentes bygger på hverdagsobservasjoner av barna i deres naturlige miljø. Dette minsker risikoen for at barnas prestasjoner blir farget av testsituasjonen.

4. Gjør testen nøyaktige målinger?

Dette spørsmålet handlet om spesifisitet og sensitivitet. Altså om testen klarer å plassere de barna som har språkvansker i riktig ”bås” og de uten vansker i riktig ”bås”. Dette handler om testens nøyaktighet i vurdering av om barna har vansker eller ikke. For at slutninger trukket fra testresultatene skal være valide, bør disse resultatene være nøyaktige. Det bør være minst mulig under- og overrapportering av vansker.

Fordelingen til barnas resultater på Språk 5-6 er tilnærmet normalfordelt, men fordelingen for ”20 spørsmål om språkferdigheter” er svært spiss. Fordelingens spissitet kan mest sannsynlig forklares ut fra at det utvalget vi var på jakt etter bare innholdt barn som en antok ikke hadde vansker, dette førte til en opphopning av barn

som skårte normalt på spørreskjemaet. Da lav skåre på skjemaet indikerer at barnet ikke har vansker, ble fordelingen svært spiss fordi de fleste barna skårte i den nedre delen av skalaen. Dette gjør at det beste statistiske målet på korrelasjonen mellom de to variablene, Pearsons R, ikke kan brukes. I stedet ble Spearmans Rho benyttet i korrelasjonsanalysen. Denne viste en signifikant sammenheng med 99% konfidensintervall. Det betyr at en kan sette stor lit til at den sammenhengen en har funnet er virkelig og ikke tilfeldig. Krysstabellen gav en oversikt over fordelingen av skårene på de to instrumentene. Det fremgår av tabellen at det var stor enighet i plasseringen av barna, men det fantes noen unntak. Hva disse unntakene skyldes er vanskelig å vite, men en mulighet er at problemer med atferd kan ha slått ut.

Nøyaktighetsanalysen viste, ut fra de data som var tilgjengelig, at testen har for dårlig sensitivitet, men adekvat spesifisitet. Den dårlige sensitiviteten kan muligens forklares av at utvalget inneholder barn som en ikke antar har vansker med språket. Dette betyr ikke at det ikke er barn i utvalget som har språkvansker, men kanskje ville analysen vist mer sensitivitet hvis en hadde inkludert en gruppe med barn som hadde språkvansker.

I noen av analysene ble utvalget delt i to etter alder. Dette ble gjort for å undersøke om barnas alder spilte en rolle. Det kan se ut til at alder spiller en viss rolle her. Barna som ble testet går siste året i barnehagen. I denne alderen har de fleste et godt språk, men de er fremdeles i en rivende utvikling. Denne utviklingen kan ha spilt en rolle for resultatene og for reliabiliteten.

6.1 Veien videre

Analysene i denne oppgaven viser at det fremdeles er mye usikkerhet rundt testens reliabilitet og validitet. En kommer stadig tilbake til ønsket om et større utvalg av barn som også omfatter en gruppe barn med påviste språkvansker. En longitudinell studie ville også være å foretrekke. Testen trenger noen justeringer og da trenger den også en ny runde med normering og validering, kanskje ville det da bli mulig å få et større og mer variert utvalg.

Når denne testen er ferdig normert og validert vil den kunne være et godt verktøy for ansatte i barnehager rundt om i Norge. Det siste året i barnehagen er det mye som skjer og det er stor forskjell i modenhet på barna når de begynner på skolen. Den nye screeningtesten kan kanskje ses på som et verktøy i overgangen barnehage- skole. Ved at dens ”storebror” Språk 6-16 allerede har blitt brukt noen år i skolen kan denne nye førskoleutgaven skape en felles plattform for lærere og barnehageansatte å snakke om barnas språklige ferdigheter på. Kommunikasjon mellom barnehage og skole kan gi barna en tryggere skolestart. Ved at de voksne møter forberedt på barnets sterke og svake sider, kan barnet møte god tilrettelegging fra første dag. Dette gjelder både de vanlige barna og de som har vansker på enkelte områder.

Litteratur

- AERA, APA, & NCME (1999). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Baddeley, A. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology*, 18(4), 362 - 365.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417-423.
- Baddeley, A. (2002). The Psychology of Memory. In A. D. Baddeley, M. D. Kopelman & B. Wilson (Eds.), *The handbook of memory disorders* (2. ed.). Chichester: Wiley.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158-173.
- Baddeley, A., & Hitch, G. (1974). Working Memory. In H. B. Gordon (Ed.), *Psychology of Learning and Motivation* (Vol. Volume 8, pp. 47-89): Academic Press.
- Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575-589.
- Baddeley, A., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*, 24(4), 490-502.
- Befring, E. (2007). *Forskingsmetode med etikk og statistikk*. Oslo: Samlaget.
- Beitchman, J. H. M. D., Wilson, B. B. S., Johnson, C. J. P. D., Atkinson, L. P. D., Young, A. P. D., Adlaf, E. P. D., et al. (2001). Fourteen-Year Follow-up of Speech/Language-Impaired and Control Children: Psychiatric Outcome. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(1), 75-82.
- Bishop, D. V. M. (1997). *Uncommon understanding: development and disorders of language comprehension in children*. Hove: Psychology Press.
- Black, J. (1979). Assessing Kindergarten Children's Communicative Competence. In O. K. Garnica & M. L. King (Eds.), *Language, Children and Society; the effect of social factors on children learning to communicate* (pp. 37 - 53). Ohio: Pergamon Press.
- Botting, N., & Conti-Ramsden, G. (2000). Social and behavioural difficulties in children with language impairment. *Child Language Teaching and Therapy*, 16(2), 105-120.
- Bronfenbrenner, U. (1979). *The ecology of human development: experiments by nature and design*. Cambridge, Mass.: Harvard University Press.
- Carey, S. (1978). The Child as word learner. In M. Hall, J. Bresnan & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264-293). Cambridge, Massachusetts: MIT Press.
- Christophersen, K.-A. (2006). *Databehandling og statistisk analyse med SPSS*. [Oslo]: Unipub.

-
- Connolly, P. (2007). *Quantitative data analysis in education: a critical introduction using SPSS*. London: Routledge.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, Tex.: Holt, Rinehart, and Winston.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3 - 17). Hillsdale, NJ: Erlbaum.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders*. Philadelphia: Saunders.
- De Vaus, D. A. (2002). *Surveys in social research*. London: Routledge.
- Dewey, D. (1995). What Is Developmental Dyspraxia. *Brain and Cognition*, 29(3), 254-274.
- Fletcher, J. M., & Satz, P. (1984). Test-Based Versus Teacher-Based Predictions of Academic Achievement: A Three-Year Longitudinal Follow-Up. *J. Pediatr. Psychol.*, 9(2), 193-203.
- Flynn, J., & Rahbar, M. H. (1998). Improving teacher prediction of children at risk for reading failure. *Psychology in the Schools*, 35(2), 163-172.
- Gall, M. D., Gall, J. P., & Borg, W. R. (1996). *Educational research: an introduction*. N.Y.: Longman.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: an introduction*. Boston, Mass.: Allyn and Bacon.
- Gathercole, S. E., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29(3), 336-360.
- Gathercole, S. E., & Baddeley, A. (1993). *Working memory and language*. Hove: Erlbaum.
- Glascoe, F. P. (1997). Parents' Concerns About Children's Development: Prescreening Technique or Screening Test? *Pediatrics*, 99(4), 522-528.
- Glascoe, F. P. (2001). Are Overreferrals on Developmental Screening Tests Really a Problem? *Arch Pediatr Adolesc Med*, 155(1), 54-59.
- Goodwin, L., D., & Leech, N., L. (2003). The meaning of validity in the new Standards for Educational and Psychological Testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development*, 36(3), 181.
- Gregory, R. J. (2000). *Psychological testing: history, principles, and applications*. Boston: Allyn and Bacon.
- Hjardemaal, F. (2002). Vitenskapsteori. In T. A. Kleven (Ed.), *Innføring i pedagogisk forskningsmetode - En hjelp til kritisk tolkning og vurdering* (pp. 28-60). Oslo: Unipub.
- Hulme, C., & Roodenrys, S. (1995). Practitioner Review: Verbal Working Memory Development and its Disorders. *Journal of Child Psychology and Psychiatry*, 36(3), 373-398.

-
- Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology*, 47(1), 72-87.
- Hundeide, K. (2008). Det intersubjektive rommet og konstruksjon av svikt og tapere. Utkast til et dialogisk og intersubjektivt alternativ. In I. Bele (Ed.), *Språkvansker; Teoretiske perspektiver og praktiske utfordringer*. Oslo: Cappelen akademisk forlag.
- Kleven, T. A. (2002a). Ikke-eksperimentelle design. In T. Lund (Ed.), *Innføring i forskningsmetodologi* (pp. 265 - 287). Oslo: Unipub.
- Kleven, T. A. (2002b). Kapittel 5: Hvordan er begrepene operasjonalisert? - Spørsmålet om begrepsvaliditet. In T. A. Kleven (Ed.), *Innføring i pedagogisk forskningsmetode - en hjelp til kritisk tolking og vurdering* (pp. 120 - 138). Oslo: Unipub.
- Knoff, H. M., Stollar, S. A., Johnson, J. J., & Chenneville, T. A. (1999). Assessment of Social- Emotional Functioning and Adaptive Behavior. In E. V. Nuttall, I. Romero & J. Kalesnik (Eds.), *Assessing and Acreening Preeschoolers; Psychological and Educational Dimentions* (pp. 126 - 161): Allyn & Bacon.
- Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, Mass.: MIT Press.
- Lichtenstein, R., & Ireton, H. (1984). *Preeschool screening; Identifying Young Children with Developmental and Educational Problems*. Orlando, Florida: Grune & Stratton, Inc.
- Marquardt, T. P., & Gillam, R. B. (1999). Assessment in communication disorders: some observations on current issues. *Language Testing*, 16(3), 249-269.
- Murphy, K. R., & Davidshofer, C. O. (2001). *Psychological testing: principles and applications*. Upper Saddle River, NJ: Prentice Hall.
- Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, 78(4, Pt.1), 679-684.
- NESH (2006). *Forskningsetiske retningslinjer for samfunnsvitenskap, jus, og teologi*. from <http://www.etikkom.no/retningslinjer/NESHretningslinjer/NESHretningslinjer/06>.
- OECD (1998). *Education at a Glance 1998: OECD Indicators*: OECD Publishing.
- Ottem, E. (2007). Språk 6- 16 : Screening-test Manual II Nye normer til bruk i forskning og utredning. *Skolepsykologi*, 4, 13 - 23.
- Ottem, E. (2009). 20 spørsmål om språkferdigheter - en analyse av sammenhengen mellom observasjonsdata og testdata. *Skolepsykologi*, 44(1), 11 - 29.
- Ottem, E., & Frost, J. (2005). *Språk 6-16, screening test, manual*. Oslo: Bredtvet kompetansesenter.
- Ottem, E., & Lian, A. (2008). Spesifikke språkvansker II: teori og empiri i kognitiv psykologisk forskning. In I. Bele & T. Helland (Eds.), *Språkvansker: teoretiske perspektiver og praktiske utfordringer* (pp. s.43-58). Oslo: Cappelen akademisk.
- Pallant, J. (2005). *SPSS survival manual: a step by step guide to data analysis using SPSS for Windows (Version 12)*. Maidenhead: Open University Press.

-
- Robson, C. (2002). *Real world research: a resource for social scientists and practitioner-researchers*. Oxford: Blackwell.
- Rydz, D., Shevell, M. I., Majnemer, A., & Oskoui, M. (2005). Topical Review: Developmental Screening. 20(1), 4-21.
- Rygvoid, A.-L. (2001). Språk- og talevansker. In T. Ogden & A.-L. Rygvold (Eds.), *Innføring i spesialpedagogikk* (pp. 197 - 231): Gyldendal Akademisk.
- Sbordone, R. J. (1996). Ecological validity: some critical issues of the neuropsychologist. . In R. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 15 - 41). Delray Beach, FL: St Lucie Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Skaalvik, E. M., & Skaalvik, S. (1996). *Selvoppfatning, motivasjon og læringsmiljø*. [Oslo]: TANO.
- Strøm, B., Borge, L.-E., & Haugsbakken, H. (2009). *Tidsbruk og organisering i grunnskolen: Sluttrapport* (No. 04/09). Trondheim: Senter for økonomisk forskning.
- Svartdal, F. (2004). *Psykologiens forskningsmetoder: en introduksjon*. Bergen: Fagbokforl.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston, Mass.: Allyn and Bacon.
- Thurlow, M. L., & Gilman, C. J. (1999). Issues and Practices in the Screening of Preeschool Children. In E. V. Nuttall, I. Romero & J. Kalesnik (Eds.), *Assessing and Screening Preeschoolers; Psychological and Educational Dimentions* (pp. 72 - 94): Allyn & Bacon
- Vedeler, L. (2000). *Observasjonsforskning i pedagogiske fag: en innføring i bruk av metoder*. Oslo: Gyldendal akademisk.
- Waters, G. S., Rochon, E., & Caplan, D. (1992). The role of high-level speech planning in rehearsal: Evidence from patients with apraxia of speech. *Journal of Memory and Language*, 31(1), 54-73.
- Wiig, E., Secord, H., & Semel, E. (1992). *CELF-Preschool. Clinical evaluation of language fundamentals-preschool. Examiner's manual*: The psychological corporation.

7. Vedlegg 1 "20 spørsmål om språkferdigheter

2

 **Bredtvet kompetansesenter**
Statlig spesialpedagogisk støttesystem

Generelle opplysninger	Fyll ut
Kjønn (1= jente, 2= gutt)	
Bostedskommune	
ID (skole/trinn/nr på klasseliste)	
Har barnet IOP? (1= ja, 2= nei, 3= vet ikke)	
Motorikk (1= bekymring, 2= normal, 3= utredet vanske, 4= vet ikke)	
Hørsel (1= bekymring, 2= normal, 3= utredet vanske, 4= vet ikke)	
Artikulasjon (1= bekymring, 2= normal, 3= utredet vanske, 4= vet ikke)	
Har perioder der han/hun stammer (1= bekymring, 2= normal, 3= utredet vanske, 4= vet ikke)	
Snakker fort, utydelig og usammenhengende (1= bekymring, 2= normal, 3= utredet vanske, 4= vet ikke)	
Syn (1= bekymring, 2= normal, 3= utredet vanske, 4= vet ikke)	
Er barnet født i Norge? (1= ja, 2= nei, 3= vet ikke)	
Hvis nei, barnets alder da det kom til Norge? (antall år)	
Barnets hovedvanske (1= ingen, 2= språkvansker, 3= lesevansker, 4= syndrom, 5= atferdsvansker, 6= annet, 7= vet ikke)	
Er barnet henvist til PPT? (1= ja, 2= nei, 3= vet ikke)	
Er det etablert et tverrfaglig/etatlig samarbeid rundt barnet (1= ja, 2= nei, 3= vet ikke)	
Generelle opplysninger om språk	
Barnets morsmål	
Mors morsmål	
Fars morsmål	
Barnets hjemmespråk	

20 spørsmål om språkferdigheter

Sett i forhold til barn på samme alder, hvor godt passer følgende utsagn for barnet? Bruk skalaen fra 1 til 5. Vennligst kryss av for hvert utsagn. Prøv å svare på alt selv om du ikke er helt sikker eller synes utsagnet virker rart.

	Helt galt			Helt riktig	
	1	2	3	4	5
<i>Semantisk evne</i>					
1. Glemmer ord som han/hun vet hva betyr	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Blander sammen ord med lik mening (for eksempel skjorte, genser, jakke)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Har vansker med å forstå hva vanlige ord betyr	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Har problemer med å svare på spørsmål like raskt som andre	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Leter ofte etter de riktige ordene	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Bruker ufullstendige setninger	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Bruker korte setninger når han/hun skal svare på spørsmål	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Har vansker med å gjenfortelle en historie han/hun har hørt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Vansker med å motta et budskap</i>					
9. Det virker som om det han/hun lærer ikke fester seg	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Har vansker med å huske ting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Vansker med å forstå hva andre sier	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Misforstår muntlige instruksjoner og beskjeder	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Har problemer med å huske beskjeder	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Misoppfatter situasjoner og hendelser	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>Vansker med å formidle et budskap</i>					
15. Eleven er vanskelig å forstå	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Problemer med å gi uttrykk for behov	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Blir ikke forstått av andre	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Tar sjelden språklige initiativ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Har problemer med å uttale ord riktig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Mestrer ikke dialog med jevnaldrende	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



Relasjoner til andre

Sett i forhold til barn på samme alder, hvor godt passer følgende utsagn for barnet? Bruk skalaen fra 1 til 5. Vennligst kryss av for hvert utsagn. Prøv å svare på alt selv om du ikke er helt sikker eller synes utsagnet virker rart.

	Helt galt			Helt riktig		
	1	2	3	4	5	
1. Er sjelden sammen med andre barn	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
2. Trenger ofte hjelp for å komme overens med andre	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
3. Andre barn blir ofte sinte på han/henne	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
4. Gir uttrykk for ubehag	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
5. Det virker som om han/hun avviser andre barn	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
6. Er ofte fortvilt og frustrert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
7. Blir ofte avvist av andre barn	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

8. Vedlegg 2: Språk 5 – 6

1

Språk 6(5)-16

av Ernst Ottem og Jørgen Frost

Kjønn: ☐ Gutt ☐ Jente ☐

Testet av:

Trinn.

Barnehage:

	År	Måned	Dag
Testdato			
Fødselsdato			
Kronologisk alder			

Aldersgruppe		Råskårer		Skalerte skårer
6 (5) til 16	Screening tester			
	Setningsminne		→	
	Ordspenn		→	
	Begreper		→	
		↓		
	Sum screening tester		→	

Screening: Minne

1. SETNINGSMINNE	
Avbryt etter tre ufullstendige setninger på rad. Ingen gjentakelser.	
Instruksjon: <i>Nå skal jeg lese noen setninger høyt for deg. Du må høre godt etter og gjenta akkurat det jeg sier.</i>	
Skåring: Notér 1 dersom setningen er helt riktig gjengitt, 0 dersom det er 1 eller flere feil	
Prøveoppgave: Kari har ei grønn kåpe	Skåre
Barn i aldersgruppen 5 til 6 år starter her:	
1. Gutten løper	
2. Mannen spiser.	
3. Jenta er glad.	
4. Hunden er stor.	
Barn som er fylt 6 år eller eldre starter her:	
5. Han sprang ut igjen.	
6. Først tapte vi, så vant vi.	
7. På skolen leker vi med bokstaver.	
8. Jenta sparket fotballen over hustaket.	
9. Mor og far reiste ut på tur i en liten kano.	
10. Pærene i hagen min er bedre enn dem i butikken.	
11. Line hadde en fin kjole på seg, fordi hun skulle være med mor.	
12. Vi vet at noen barn slutter å gråte når vi gir dem noe å spise.	
13. Hun er så flink til å spille gitar at hun kan bli med i et rockeband.	
14. Jenta stoppet og kjøpte brus, selv om hun kom for sent på skolen.	
15. Hvis det ikke slutter å regne snart, blir de nødt til å avlyse utekonserten.	
16. Før elevene på videregående fikk gå ut, måtte de levere inn besvarelsene på matteprøvene.	
17. Gutten som skulka treninga, fikk ikke lov til å spille på laget før det var gått en uke	
18. Dersom treneren hadde latt oss slutte treninga tidligere, ville vi ha vært hjemme for lenge siden.	
19. Da skoledagen var slutt, bestemte elevene seg for å sykle en tur før de dro hjem.	
20. Bibliotekaren har tolv nye mattebøker for åttende klasse, som er reservert for oss.	
Sum	

2. ORDSPENN	
Avbryt etter tre nullskårer på rad. Ingen gjentakelser.	
Instruksjon: <i>Nå skal jeg si noen ord etter hverandre og du skal si de samme ordene.</i> Ordene leses med en hastighet på ca. ett ord pr. sekund. Skriv ned det barnet sier i samme rekkefølge som det barnet sier.	
Skåring: Notér 1 dersom alle ord er i riktig rekkefølge, 0 dersom ett eller flere av ordene er utelatt, innskutt (nytt ord) eller i feil rekkefølge.	
Prøveoppgave: Tann – Sekk – Kokk	Skåre
To ord: Barn i aldersgruppen 5 til 6 år starter her:	
1. Regn – Grønn	
2. Drikk – Fikk	
3. Ball – Rusk	
4. Tann – Land	
Tre ord: Barn som er fylt 6 år eller eldre starter her:	
5. Sønn – Tull – Klegg	
6. Sild – Mild – Vill	
7. Rusk – Hegg – Fikk	
8. Ball – Kall – Fall	
Fire ord	
9. Sann – Rygg – Kokk – Mutt	
10. Tinn – Vind – Finn – Kinn	
11. Grønn – Kan – Kutt – Vind	
12. Tann – Kan – Sann – Land	
Fem ord	
13. Brekk – Lett – Kull – Rygg – Tann	
14. Drønn – Gutt – Flekk – Takk – Dregg	
15. Hegg – Legg – Vegg – Dregg – Klegg	
16. Butt – Kutt – Mutt – Putt – Gutt	
Sum	

Screening: Begreper

3. MOTSETNINGER		
Avbryt etter tre feil svar etter hverandre. Gjentakelser tillatt.		
Instruksjon: Nå skal du finne det motsatte ordet av det jeg sier.		
Skåring: Notér 1 for riktig svar, 0 for feil svar.		
Prøveoppgave: Er kapteinen liten eller er han _____ (Stor)?		Skåre
Alle barn starter her:		
1. Er jakken pen eller er den _____ (Stygg)?		
2. Er gulvet vått eller er det _____ (Tørt)?		
3. Er buksene vide eller er de _____ (Trange, smale)?		
Hva er det motsatte av:		
4. Starte	(Stoppe, slutte)	
5. Feig	(Tapper, modig, tøff)	
6. Tam	(Vill)	
7. Synke	(Flyte, stige)	
8. Krympe	(Vokse)	
9. Sende	(Ta i mot, motta, få)	
10. Stramme	(Løse, slakke)	
11. Finne	(Miste, gjemme)	
12. Tømme	(Fylle)	
13. Optimist	(Pessimist)	
14. Konkret	(Abstrakt)	

4. ORDKUNNSKAP	
Avbryt etter tre nullskårer på rad. Gjentakelser tillatt.	
Instruksjon: <i>Hva er ...?</i> eller <i>Hva betyr...?</i> Skåring: Notér 1 for alle betydninger av et ord som man kan finne i vanlige ordbøker. 0 dersom svaret er åpenbart galt, svært vagt eller har et fattig meningsinnhold (jf. Appendiks i manualen)	
Prøveoppgave: Hva er en hatt? (Til å ha på hodet, et hodeplagg)	
Barn i aldersgruppen 5 til 6 år starter her:	Skåre
1. Hva er en katt?	
2. Hva er en gave?	
3. Hva er en telefon?	
Barn som er fylt 6 år eller eldre starter her:	
4. Hva er en paraply? (Den beskytter deg mot regn)	
5. Hva er en seng? (Sted å sove, møbel, madrass)	
6. Hva er en tulipan? (Plante, blomst, den vokser og gror)	
7. Hva er et brev? (Skriver og sender, et kort, legger i postkassen)	
8. Hva er en spiker? (Banke, hamre den inn i noe. Den er skarp og til å banke på)	
9. Hva er en belønning? (Finnerlønn, dusør, erstatning, vederlag, påskjønnelse, noe du får når du har gjort noe)	
10. Hva er et mikroskop? (Et instrument du kan se små ting i)	
11. Hva betyr presis? (Punktlig, nøyaktig, akkurat)	
12. Hva betyr lojalitet? (Trofasthet, støtte, troskap)	
13. Hva er en fabel? (Historie, myte, legende)	
14. Hva er en fure? (Fordypning, rynke)	
15. Hva betyr konsumere? (Fortære, forbruke)	
Sum for deltestene 3 + 4	

9. Vedlegg nr 3: Informasjonsbrev til foreldrene

Til foreldre/foresatte ved _____ barnehage.

Oslo, 11.12.08

Foreldretillatelse

I regi av Bredtvet kompetansesenter og Institutt for spesialpedagogikk, Universitetet i Oslo er det utviklet en ny språkscreeningstest for 5-åringer. For å normere denne trenger vi å prøve denne ut på mange 5-åringer uten antatte språkvansker.

Vi ber i denne forbindelse om tillatelse til å prøve ulike språktester på ditt barn. Disse måler minne, språkforståelse, ordforråd og grammatikk. I tillegg vil barnets førskolelærer fylle ut et skjema med spørsmål om barnets språklige fungering. Gjennomføringen vil ta ca 45 minutter og vil bli delt med pauser hvis barnet har behov for det. Utprøvingen vil skje i barnehagen, og vil bli tilpasset barnehagens daglige rutiner i perioden januar-februar 2009. Testene vil foregå under trygge forhold og vil etter vår mening oppleves som interessant og spennende for barnet å delta i.

Fem masterstudenter; Ingeborg Watne, Ida Krohn, Hege Mari Sørli, Monica Lillevold og Mona B. Heier, ved Institutt for spesialpedagogikk (ISP), Universitetet i Oslo, vil lede testene og skrive sine masteroppgaver i tilknytning til prosjektet. Masterstudenten har taushetsplikt, og prosjektet er meldt til datatilsynet.

De opplysningene vi trenger i tillegg er barnets kjønn og fødselsdato.

Det er selvsagt frivillig å delta, men det vil være ønskelig for oss at flest mulige 5 åringer deltar i dette opplegget. Alle besvarelser vil bli anonymisert, det vil si at det ikke er mulig å gjenkjenne det enkelte barn.

På enkelte deltester vil det bli benyttet lydopptak. Grunnen til dette er at vi skal sikre oss å få med alt det barnet sier. Lydopptakene vil ikke bli lagret på PC. Lydopptakene vil bli slettet umiddelbart etter at barnas uttalelser er skrevet ned.

Alle testprotokoller vil bli anonymisert umiddelbart etter testing i januar/februar. Barna vil bli registrert med tall og fødselsdato. Barnehagen vil tildele barna et nummer. Listen som kobler barnets navn til nummer vil bli oppbevart i barnehagen, men både listen og testprotokollene vil bli makulert ved prosjektslutt i august 2009.

Testresultatene skal brukes ved UiO og ved Bredtvet som anonym statistikk.

I tillegg til språktestene skal ped.leder gi noen opplysninger om barnets spåkferdigheter og relasjoner til andre i et observasjonsskjema.

Det er mulighet til å trekke seg fra deltagelse fram til opplysningene er anonymisert.

Hvis du/dere skulle ha noen spørsmål, vennligst ta kontakt med undertegnede

Hvis du/dere ønsker at ditt/deres barn skal delta, vær vennlig å underskrive svarslippen og returner den til barnehagen innen.....

Med vennlig hilsen

Marit Holm
Institutt for spesialpedagogikk

Jan Mossige
Bredtvet kompetansesenter

Student som skal lede testene i din barnehage: Monica Lillevold

e-post:

tlf.